



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number 09/761466

TO: Celine Qian
Location: CM1/11C10/11E12
Art Unit : 1636
Thursday, August 21, 2003

Case Serial Number: 09/761466

From: Mary Jane Ruhl
Location: Biotech-Chem Library
CM1-6A06
Phone: 605-1155

Maryjane.ruhl@uspto.gov

Search Notes

Examiner Qian,

Here are the results from your search.

Please let me know if you have any questions or need further assistance.

Sincerely,
Mary Jane Ruhl



BEST AVAILABLE COPY


```

;
; COUNTRY: USA
; ZIP: 08053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,982A
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/756,806
; FILING DATE: No. 5981731ember 26, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07111
; FILING DATE: May 31, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/250,856
; FILING DATE: May 31, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISPH-0212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 779-8488
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2510
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
; US-08-888-982A-42

```

```

Query Match 7.4%; Score 35.2; DB 2; Length 2510;
Best Local Similarity 50.6%; Pred. No. 0.86;
Matches 85; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

```

OY 88 CTTTGTGTACACCTGACCCACAGAGCTGACGTTCTAGAGAGCCCTTACCGAATAAT 147
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2343 CGTTTCTGTCCACTGAACAAATGAGTGAGAGAGTTGACAGAGTGCACAAAGAA 2402
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 148 AGGAATAATTCCTGACCTTGATTGCAAGGCAATGCTAATTTTCTTCTCA 207
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2403 AATAAATGAACATATGTTGCTATATGTAATAATTAATAATGCTCTTTT 2462
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 208 GAGCTCTCAAAAAAAAAAAAAAACTTACTTAAAAACAGGATCC 255
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2463 AGGTGGAATAAAAAAAAAAAAAAAACCC 2510
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 3
US-09-462-261-42
; Sequence 42, Application US/09462261
; Patent No. 6391636
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; TITLE OF INVENTION: Antisense Oligonucleotide
; Modulation of raf Gene Expression
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 66 East Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: Pentium
; OPERATING SYSTEM: Windows 95

```

```

;
; SOFTWARE: WORDPERFECT 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/462,261
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/756,806
; FILING DATE: No. 6391636ember 26, 1996
; APPLICATION NUMBER: PCT/US95/07111
; FILING DATE: May 31, 1995
; APPLICATION NUMBER: 08/250,856
; FILING DATE: May 31, 1994
; APPLICATION NUMBER: 08/888,982
; FILING DATE: July 7, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISPH-0312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 810-1454
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2510
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
; US-09-462-261-42

```

```

Query Match 7.4%; Score 35.2; DB 4; Length 2510;
Best Local Similarity 50.6%; Pred. No. 0.86;
Matches 85; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

```

OY 88 CTTTGTGTAAACCTGACCCACAGAGCTGACGTTCTAGAGAGCCCTTACCGAATAAT 147
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2343 CGTTTCTGTCCACTGAACAAATGAGTGAGAGAGTTGACAGAGTGCACAAAGAA 2402
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 148 AGGAATAATTCCTGACCTTGATTGCAAGGCAATGCTAATTTTCTTCTCA 207
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2403 AATAAATGAACATATGTTGCTTATGTTAATTAATAATGCTCTTTT 2462
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 208 GAGCTCTCAAAAAAAAAAAAAAACTTACTTAAAAACAGGATCC 255
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2463 AGGTGGAATAAAAAAAAAAAAAAAACCC 2510
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 4
US-09-506-073-89
; Sequence 89, Application US/09506073
; Patent No. 6410518
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; TITLE OF INVENTION: Antisense Oligonucleotide Modulation of raf Gene Expression
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/506,073
; CURRENT FILING DATE: 2000-02-18
; EARLIER APPLICATION NUMBER: US 09/143,214
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: PCT/US98/13961
; EARLIER FILING DATE: 1998-07-06
; EARLIER APPLICATION NUMBER: US 08/888,982
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US 08/756,806
; EARLIER FILING DATE: 1996-11-26
; EARLIER APPLICATION NUMBER: PCT/US95/07111
; EARLIER FILING DATE: 1995-05-31
; EARLIER APPLICATION NUMBER: US 08/250,856
; EARLIER FILING DATE: 1994-05-31
; NUMBER OF SEQ ID NOS: 130
; SEQ ID NO 89
; LENGTH: 2510

```


[illegible]

RESULT 5
 OS-08-700-026, 2
 SUBJECT 2 APPLICABLE DS/ 07 of 2
 Patent No. 6,743,497
 GENERAL INFORMATION:
 APPLICANT: Asystech, Ltd.
 APPLICANT: Alameda, California
 APPLICANT: 1411 S. 17th St.
 TITLE OF INVENTION: NOVEL HUMAN IRI/ACHP-LIKE PROTEIN
 NUMBER OF SCIENTISTS: 1
 CORRESPONDENT ADDRESS:
 ADDRESSEE: Lucy Pharmaceuticals, Inc.
 STREET: 4174 Barrett Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: U.S.
 ZIP: 94304
 MOTION PICTURE FILE NO.
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSS, VORSEAL 1.1
 CURRENT APPLICANT DATA:
 APPLICATION NUMBER: OS/08/700-026
 FILING DATE: Filed 10/05/97
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 96,744
 REFERENCE/DOCKET NUMBER: 96-0115 DS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4176
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 114 base pairs
 TYPE: nucleotide
 STRANDEDNESS: single
 topolgy: +sense
 METHOD OF DETERMINATION: N/A
 IMMEDIATE SOURCE:
 LIBRARY: Plasmid
 SOURCE: 627904
 OS-08-700-026, 2
 Query Match 74% Score 45; DB J: Genbank 1145
 Host Local: Simularity 49.9%; Local No. 9712
 Matches 51; Unscored 100; Mismatches 90; Indels 0; Caps 0
 114 AATAAATTTTCTTAAATTAATTAATTAATTAATTAATTTTCTTTTCTAAATG 211
 1-45 AATAAATTTTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTTCTTTTCTAAATG 211

27	z12	1010AAAAAAAAAAAAAA	2-4
1ab	1095	1111111111111111	
		100000000000000000	11.7

BEST 1 5
 25-03-17 4-000-4
 SOURCE: 5, APPLICANT: US/921845
 EXAMINER: 6451361
 GENERAL INFORMATION:
 APPLICANT: Farrow, Steven Earl
 APPLICANT: Hitz, William D.
 APPLICANT: Kinney, Anthony J.
 APPLICANT: Kuroda, Kazuo
 APPLICANT: Katsuki, J. Acon
 TITLE OF INVENTION: PLANT ARCHITECTURAL MAIN ARM/ WITH RESIN SHEET
 FILE REFERENCE: 66 1126
 CURRENT APPLICATION NUMBER: US/93/174,306
 CURRENT FILING DATE: 1998-10-15
 EARLIER APPLICATION NUMBER: 66/064,428
 EARLIER FILING DATE: 1987 October 20
 NUMBER OF SEQ ID NOS: 54
 SOFTWARE: Microsoft Word Version 7.0A
 SEQ ID NO: 3
 LENGTH: 2078
 FILE: LNA
 APPLICANT: Olivevine max
 25-03-17 4-000-4

[illegible]

```

PSID: 7
Seq-ID No: 9218
Source: Applicant: US/94187218
Patent No.: 607687
GENERAL INFORMATION:
APPLICANT: Leiden Bioscience Z.
APPELLANT: Badzysowski, Gregory J.
APPELLOAN: Father, Sharon E.
APPELLANT: Wachtel, Lynette M.
TITLE OF INVENTION: Herbicide Target Sites and Methods
FILE PERCENTAGE: 89/5 40780A
CURRENT APPLICATION NUMBER: US/94/9801 9218
CURRENT FILING DATE: 2000-01-11
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patcom V.1. 2.1
SEQ ID NO: 18
LENGTH: 218
FEEDBACK:
REMARKS: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_109718
1 CALL ME: (1) : (718)
OTHER INFORMATION: a, b, c, d, e, f
US-94-180 9218 18

```

[illegible]

```
Db      161 COTTATAAATTATTATTTCATTAGCCCAAAAAAAAAAAAAAAAAAAAA 214
RESULT 8
US-09-149-476-191
; Sequence 191, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; EARLIER FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
```

```
1 EARLIER APPLICATION NUMBER: 60/057,761
2 EARLIER FILING DATE: 1997-08-22
3 EARLIER APPLICATION NUMBER: 60/047,515
4 EARLIER FILING DATE: 1997-05-24
5 EARLIER APPLICATION NUMBER: 6/02047,599
6 EARLIER FILING DATE: 1997-05-24
7 EARLIER APPLICATION NUMBER: 60/047,598
8 EARLIER FILING DATE: 1997-05-24
9 EARLIER APPLICATION NUMBER: 6/047,585
10 EARLIER FILING DATE: 1997-05-24
11 EARLIER APPLICATION NUMBER: 60/047,586
12 EARLIER FILING DATE: 1997-05-24
13 EARLIER APPLICATION NUMBER: 60/047,590
14 EARLIER FILING DATE: 1997-05-23
15 EARLIER APPLICATION NUMBER: 60/047,594
16 EARLIER FILING DATE: 1997-05-23
17 EARLIER APPLICATION NUMBER: 60/047,589
18 EARLIER FILING DATE: 1997-05-23
19 EARLIER APPLICATION NUMBER: 60/047,593
20 EARLIER FILING DATE: 1997-05-23
21 EARLIER APPLICATION NUMBER: 60/047,614
22 EARLIER FILING DATE: 1997-05-23
23 EARLIER APPLICATION NUMBER: 60/044,578
24 EARLIER FILING DATE: 1997-04-11
25 EARLIER APPLICATION NUMBER: 60/043,576
26 EARLIER FILING DATE: 1997-04-11
27 EARLIER APPLICATION NUMBER: 60/047,501
28 EARLIER FILING DATE: 1997-05-22
29 EARLIER APPLICATION NUMBER: 60/043,570
30 EARLIER FILING DATE: 1997-04-11
31 EARLIER APPLICATION NUMBER: 60/056,512
32 EARLIER FILING DATE: 1997-08-22
33 EARLIER APPLICATION NUMBER: 6/056,564
34 EARLIER FILING DATE: 1997-08-22
35 EARLIER APPLICATION NUMBER: 60/056,876
36 EARLIER FILING DATE: 1997-08-22
37 EARLIER APPLICATION NUMBER: 6/056,881
38 EARLIER FILING DATE: 1997-08-22
39 EARLIER APPLICATION NUMBER: 60/056,909
40 EARLIER FILING DATE: 1997-08-22
41 EARLIER APPLICATION NUMBER: 60/056,875
42 EARLIER FILING DATE: 1997-08-22
43 EARLIER APPLICATION NUMBER: 60/056,862
44 EARLIER FILING DATE: 1997-08-22
45 EARLIER APPLICATION NUMBER: 60/056,867
46 EARLIER FILING DATE: 1997-08-22
47 EARLIER APPLICATION NUMBER: 60/056,908
48 EARLIER FILING DATE: 1997-08-22
49 EARLIER APPLICATION NUMBER: 60/048,964
50 EARLIER FILING DATE: 1997-08-22
51 EARLIER APPLICATION NUMBER: 60/057,650
52 EARLIER FILING DATE: 1997-04-05
53 EARLIER APPLICATION NUMBER: 60/056,884
54 EARLIER FILING DATE: 1997-08-22
55 EARLIER APPLICATION NUMBER: 60/057,669
56 EARLIER FILING DATE: 1997-09-05
57 EARLIER APPLICATION NUMBER: 60/049,610
58 EARLIER FILING DATE: 1997-09-13
59 EARLIER APPLICATION NUMBER: 60/061,060
60 EARLIER FILING DATE: 1997-10-02
61
62 Query Match: 7.98; State 44.6; Db 4; Length 2779;
63 Best Local Similarity: 61.48; Freq. No. 2.0;
64 Matches: 54; Conservative: 0; Mismatches: 34; Indels: 0; Gaps: 0;
65
66 193 GGTACTTGATTTGCAAGGTAATATTAATTTTCTTCTCGAGAGCTGTCAAAAAA 222
67 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
68 2691 GGGCAGGGAATTTGTTTCTTAATATTTGATCTACTGCAAAAAA 2750
69 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
70 223 AAAAAAAGCTTACTTAAAAAGG 250
71 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
72 2751 AAAAAAAGCTTAAAAAAGG 2778
73 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
685011 9
685012 40, Application US/09227,457
685013 Patent No. 6342581
685014 GENBANK INFORMATION:
685015 TITLE OF INVENTION: 123 Human secreted proteins
685016
685017 FILE REFERENCE: P20101
685018 CURRENT FILING DATE: 1999-01-08
685019 EARLIER APPLICATION NUMBER: P51/09298/13664
685020 EARLIER FILING DATE: 1998-07-07
685021 EARLIER APPLICATION NUMBER: 60/051,426
685022 EARLIER FILING DATE: 1997-07-08
685023 EARLIER APPLICATION NUMBER: 60/052,732
685024 EARLIER FILING DATE: 1997-07-08
685025 EARLIER APPLICATION NUMBER: 60/051,945
685026 EARLIER FILING DATE: 1997-07-08
685027 EARLIER APPLICATION NUMBER: 60/051,929
685028 EARLIER FILING DATE: 1997-07-08
685029 EARLIER APPLICATION NUMBER: 60/052,803
685030 EARLIER FILING DATE: 1997-07-08
685031 EARLIER APPLICATION NUMBER: 60/052,732
685032 EARLIER FILING DATE: 1997-07-08
685033 EARLIER APPLICATION NUMBER: 60/051,932
685034 EARLIER FILING DATE: 1997-07-08
685035 EARLIER APPLICATION NUMBER: 60/051,916
685036 EARLIER FILING DATE: 1997-07-08
685037 EARLIER APPLICATION NUMBER: 60/051,920
685038 EARLIER FILING DATE: 1997-07-08
685039 EARLIER APPLICATION NUMBER: 60/052,733
685040 EARLIER FILING DATE: 1997-07-08
685041 EARLIER APPLICATION NUMBER: 60/052,795
685042 EARLIER FILING DATE: 1997-07-08
685043 EARLIER APPLICATION NUMBER: 60/051,919
685044 EARLIER FILING DATE: 1997-07-08
685045 EARLIER APPLICATION NUMBER: 60/051,928
685046 EARLIER FILING DATE: 1997-07-08
685047 EARLIER APPLICATION NUMBER: 60/055,722
685048 EARLIER FILING DATE: 1997-08-18
685049 EARLIER APPLICATION NUMBER: 60/055,723
685050 EARLIER FILING DATE: 1997-08-18
685051 EARLIER APPLICATION NUMBER: 60/055,948
685052 EARLIER FILING DATE: 1997-08-18
685053 EARLIER APPLICATION NUMBER: 60/055,949
685054 EARLIER FILING DATE: 1997-08-18
685055 EARLIER APPLICATION NUMBER: 60/055,953
685056 EARLIER FILING DATE: 1997-08-18
685057 EARLIER APPLICATION NUMBER: 60/055,950
685058 EARLIER FILING DATE: 1997-08-18
685059 EARLIER APPLICATION NUMBER: 60/055,947
685060 EARLIER FILING DATE: 1997-08-18
685061 EARLIER APPLICATION NUMBER: 60/055,964
685062 EARLIER FILING DATE: 1997-08-18
685063 EARLIER APPLICATION NUMBER: 60/056,960
685064 EARLIER FILING DATE: 1997-09-18
685065 EARLIER APPLICATION NUMBER: 60/056,984
685066 EARLIER FILING DATE: 1997-09-18
685067 EARLIER APPLICATION NUMBER: 60/056,983
685068 EARLIER FILING DATE: 1997-08-18
685069 EARLIER APPLICATION NUMBER: 60/056,983
685070 EARLIER FILING DATE: 1997-08-18
685071 EARLIER APPLICATION NUMBER: 60/056,983
685072 EARLIER FILING DATE: 1997-08-18
685073 EARLIER APPLICATION NUMBER: 60/056,983
685074 EARLIER FILING DATE: 1997-08-18
685075 EARLIER APPLICATION NUMBER: 60/056,983
685076 EARLIER FILING DATE: 1997-08-18
685077 EARLIER APPLICATION NUMBER: 60/056,983
685078 EARLIER FILING DATE: 1997-08-18
685079 EARLIER APPLICATION NUMBER: 60/056,983
685080 EARLIER FILING DATE: 1997-08-18
685081 EARLIER APPLICATION NUMBER: 60/056,983
685082 EARLIER FILING DATE: 1997-08-18
685083 EARLIER APPLICATION NUMBER: 60/056,983
685084 EARLIER FILING DATE: 1997-08-18
685085 EARLIER APPLICATION NUMBER: 60/056,983
685086 EARLIER FILING DATE: 1997-08-18
685087 EARLIER APPLICATION NUMBER: 60/056,983
685088 EARLIER FILING DATE: 1997-08-18
685089 EARLIER APPLICATION NUMBER: 60/056,983
685090 EARLIER FILING DATE: 1997-08-18
685091 EARLIER APPLICATION NUMBER: 60/056,983
685092 EARLIER FILING DATE: 1997-08-18
685093 EARLIER APPLICATION NUMBER: 60/056,983
685094 EARLIER FILING DATE: 1997-08-18
685095 EARLIER APPLICATION NUMBER: 60/056,983
685096 EARLIER FILING DATE: 1997-08-18
685097 EARLIER APPLICATION NUMBER: 60/056,983
685098 EARLIER FILING DATE: 1997-08-18
685099 EARLIER APPLICATION NUMBER: 60/056,983
685100 EARLIER FILING DATE: 1997-08-18
```



```

PCT-US94-14073-1
Sequence 1, Application PC/TUS9414073
GENERAL INFORMATION:
APPLICANT: THOMAS, Wayne R., CHUA, Kaw-Yan, ROGERS, Bruce L., and
APPLICANT: KUO, Mei-chang
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING A HOUSE DUST MITE
TITLE OF INVENTION: ALLERGEN, DER P IIT, AND USES THEREFOR
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14073
FILING DATE: 08-DEC-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-072 (IMI-041)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1059 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 63..848
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 150..848
PCT-US94-14073-1

Query Match          6.9%; Score 33; DB 5; Length 1059;
Best Local Similarity 57.1%; Pred.No. 2.7;
Matches    60; Conservative   0; Mismatches   45; Indels     0; Gaps      0;

QY      144 AAATAGCAATAAATCCCTTGGCACCTTGATTTGCAAGGCGAATGCTAAATTTTTTCCTTC 203
        ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       951 ATATTCCAGCTGAAGAACCAACAAATCATGTATGATAGGAGATTGTCTGTTTCAAT 1010
        ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      204 TCCAGAGCTCTCAAAAAAAAAAAAAAAAAAAAAACCCTACTCAAAAACA 248
        ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1011 TAAAAAATCCAAAATAAATAAATAAATAAATCAATTCACAAAAA 1055

```

Search completed: August 20, 2003, 14:32:46
Job time : 51.3482 secs

PT	induction, e.g. for optimizing cardiomyocyte induction	-
PS	Claim 21; Fig 5C; 66pp; English.	
XX	The sequence represents the coding sequence of cardiac enhancer	
CC	hcsx/Nkx2.5 homology domain B. The nucleic acid is useful for	
CC	specifically expressing a gene in a cardiac cell, as an earlier marker	
CC	of cardiomyocyte induction, e.g. for optimizing cardiomyocyte induction.	
CC	Genes expressed in the cardiac cell-specific manner are useful for the	
CC	targeted expression of genes encoding therapeutic proteins for the	
CC	treatment of damaged heart tissue. Cardiac specific enhancer elements may	
CC	be used for gene therapy.	
XX	Sequence 478 BP; 138 A; 129 C; 89 G; 122 T; 0 other;	
SO	Query Match	100.0%; Score 478; DB 22; Length 478;
	Best Local Similarity	100.0%; Pred. No. 2, 2e-122;
	Matches 478; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1	AGAGAAATCATTTACCCGATTCACAAAGAGCATAGAGAGTGAACAGTCACATCTGT 60
DB	1	AGAGAAATCATTTACCCGATTCACAAAGAGCATAGAGAGTGAACAGTCACATCTGT 60
QY	61	CAATATAGGACAGTTTTTTTCTCCTTCTTTTGTAAACACCTGACCCACAGAGCATGACAG 120
DB	61	CAATATAGGACAGTTTTTTTCTCCTTCTTTTGTAAACACCTGACCCACAGAGCATGACAG 120
QY	121	TTCTAGGAAAGCCCTTACCCGAAATAGAAATAATCCTTGCCACCTTGATTTGCAAG 180
DB	121	TTCTAGGAAAGCCCTTACCCGAAATAGAAATAATCCTTGCCACCTTGATTTGCAAG 180
QY	181	GGCAATGCTAATTTTTTTTCTTCTTCACAGCTCTCAAAAAAAAAAAAAAAAAACCTTAC 240
DB	181	GGCAATGCTAATTTTTTTTCTTCTTCACAGCTCTCAAAAAAAAAAAAAAAAAACCTTAC 240
QY	241	TAAAAACGGGATCCCGATGAGCTGAGTGTCCGCCATTTAAACGGTAATTTTCAGGC 300
DB	241	TAAAAACGGGATCCCGATGAGCTGAGTGTCCGCCATTTAAACGGTAATTTTCAGGC 300
QY	301	GTCGCGTACACTAATCTTTTCAAACTGTATCGCGAGCCGCTGGCCAGACATTCACTT 360
DB	301	GTCGCGTACACTAATCTTTTCAAACTGTATCGCGAGCCGCTGGCCAGACATTCACTT 360
QY	361	AACAGCGCTCCAGGACCGCTGTCGAGCTCTTTTCACGAGACATTAATTTGAATGCG 420
DB	361	AACAGCGCTCCAGGACCGCTGTCGAGCTCTTTTCACGAGACATTAATTTGAATGCG 420
QY	421	ATGTGGCTGTTTTGGCAGACGTACCGGCTCGGCGATAGGCAATCCTCTCCAAAGACAC 478
DB	421	ATGTGGCTGTTTTGGCAGACGTACCGGCTCGGCGATAGGCAATCCTCTCCAAAGACAC 478
RESULT 2		
ASD	AA509962	
AC	AA509962	standard; DNA; 6751 BP.
XX	AA509962;	
XX	24-OCT-2001	(first entry)
DE	Genomic DNA #2 encoding human Csx/Nkx2.5.	
XX	Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;	
KW	therapeutic; heart tissue; gene therapy; human; ds.	
OS	Homo sapiens.	
PN	WO200151006-A2.	
XX	19-JUL-2001.	
PF	16-JAN-2001; 2001WO-US01511.	

PR	14-JAN-2000; 2000US-0176419.
XX	
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX	
PI	Lee IW, Izumo S;
XX	
DR	WPI: 2001-451809/48.
XX	
XX	New cardiac specific cell enhancer elements, useful for specifically
PT	expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT	induction, e.g. for optimizing cardiomyocyte induction
XX	
PS	Claim 29; Fig 4B; 66pp; English.
XX	
CC	The sequence represents the genomic sequence #2 of human Csx/Nkx2.5. The
CC	nucleic acid is useful for specifically expressing a gene in a cardiac
CC	cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC	optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC	specific manner are useful for the targeted expression of genes encoding
CC	therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC	specific enhancer elements may be used for gene therapy.
XX	
SO	Sequence 6751 BP; 1459 A; 2008 C; 1753 G; 1490 T; 41 other;
	Query Match 100.0%; Score 478; DB 22; Length 6751;
	Best Local Similarity 100.0%; Prid. No. 6e-122;
	Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 AGAGAATTCATTATCCGATTCACAAAGAGCATAGAGAGTGAACAGTCACTGATCTTGT 60
DB	201 AGAATAATTCATTATCCGATTCACAAAGAGCATAGAGAGTGAACAGTCACTGATCTTGT 260
OY	61 CAATTAGGAGAGATTTTTTTCCTTCCTTTTGTACACCTGACCCAGAGACTGACAG 120
DB	261 CAATTAGGAGAGATTTTTTTCCTTCCTTTTGTACACCTGACCCAGAGACTGACAG 320
OY	121 TTTCAGGAGAGGCCCCCTTACCCGAAATAGAAATAAATCTTGGCACCTGATTGGCAAG 180
DB	321 TTTCAGGAGAGGCCCCCTTACCCGAAATAGAAATAAATCTTGGCACCTGATTGGCAAG 380
OY	181 GGCATGCTAAATTTTTTTCCTTCCTTCACAGGCTCAAAAAAAAAAAAAAACTTAC 240
DB	381 GGCATGCTAAATTTTTTTCCTTCCTTCACAGGCTCAAAAAAAAAAAAAAACTTAC 440
OY	241 TAAATAAGAGGATCCCGGATGTAGCTGTGATGTCCTCCCATTAAGCGTAATATTACAGC 300
DB	441 TAAATAAGAGGATCCCGGATGTAGCTGTGATGTCCTCCCATTAAGCGTAATATTACAGC 500
OY	301 GTCCGCTCACACTATCTTTCCTCAACGTGCATCGGAGAGCCGCTGGCCAGCATTCACCT 360
DB	501 GTCCGCTCACACTATCTTTCCTCAACGTGCATCGGAGAGCCGCTGGCCAGCATTCACCT 560
OY	361 AACAGCGCTCCAGAGACCTCTGTTCCGAGCTCTTTTACGCGAGACATTTAATTGAATCG 420
DB	561 AACAGCGCTCCAGAGACCTCTGTTCCGAGCTCTTTTACGCGAGACATTTAATTGAATCG 620
OY	421 ATGGGCGCTGTTTCCAGAGCTCAACGCGCTGGGAGATGGGATGCTCTCCAAAGCAAC 478
DB	621 ATGGGCGCTGTTTCCAGAGCTCAACGCGCTGGGAGATGGGATGCTCTCTCCAAAGCAAC 678
RESULT 3	
AAC59534	standard; cDNA; 1086 BP.
AAC59534;	
15-FEB-2001	(first entry)
Human secreted protein cDNA sequence #28.	
Cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;	
antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;	

[illegible][illegible]

```
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-457723/49.
XX
XX DR Isolated polypeptide for treating, preventing and/or prognosing
XX PT respiratory disorders related to the lung including lung cancers and
XX PT also for testing and detection e.g. diagnosis -
XX
PS Claim 1; SEQ ID No 363; 507pp; English.

CC Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode
CC the lung antigen polypeptides of the invention. Lung antigen polypeptides
CC and their associated polynucleotides are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by detecting the presence or
CC absence of a mutation in a lung antigen polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Query Match 8.2%; Score 39; DB 22; Length 3273;
Best Local Similarity 68.4%; Pred. No. 1.3;
Matches 54; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 177 CAGGGCAATGCTAATTTTCTTTCCTCCACAGCTCTCAAAAAAAAAAAAAAAAAACC 236
DB 2059 CATGGACATGTAATTTTATTCCTCTACCTGTTTAAAAAACATRAAAAAAAAAACC 2118
QY 237 TTACTAAAAACAGGATCC 255
DB 2119 TAAAAAATAAAAAAAGTCC 2137

RESULT 5
AAS30098 ID AAS30098 standard; DNA; 3276 BP.
XX
XX AC AAS30098;
XX
XX DT 21-NOV-2001 (first entry)
XX
XX DE Human lung antigen genomic DNA #168.
XX
XX KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
```


PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241825.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.

PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0234097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.
DR P-PSDB; AAU18622.
XX
PT Isolated polypeptide for treating, preventing and/or prognosing
PT respiratory disorders related to the lung including lung cancers and
PT also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID No 81; 507pp; English.
XX
CC Sequences AAS29839-AAS29930 represent cDNA molecules, which encode the
CC lung antigen polypeptides of the invention. Lung antigen polypeptides and
CC their associated polynucleotides are useful in the diagnosis, treatment
CC and prevention of various types of disorders in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological
CC condition can be determined by detecting the presence or absence of a
CC mutation in a lung antigen polynucleotide. The treatable disorders
CC include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match 8.0%; Score 38.4; DB 22; Length 498;
Best Local Similarity 67.5%; Pred. No. 0.97;
Matches 54; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 177 CAAGGCAATGCTAATTTTCTTCCTCCACAGCTCTCAAAAAAAAAAAAAAAAAACC 236
||| ||||| ||||| ||||| ||| | ||||| ||||| ||||| |||||
Db 412 CATGGACAAATGTAATTTTATCTCTACCTGTTAAAAAACTRAAAAAAAAAACC 471

OY 237 TTACTAAAAACAGGATGCC 256
| | ||||| | | ||
Db 472 TAAAAAAAAAAAAAAAAAGTCC 491

RESULT 9
AAF94842/c
ID AAF94842 strand: cDNA; 396 BP.
XX
AC AAF94842;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human ovarian cancer associated coding sequence SEQ ID NO: 33.
XX
KW Human, ovarian cancer; vaccine; gene therapy; carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200118046-A2.
XX
PD 15-MAR-2001.
XX

[illegible]

52	proteins and polynucleotides may be used to prevent,	
53	diagnose and treat immune/hematopoietic-related diseases, especially	
54	cancers and cancer metastases of hematopoietic-derived cells. AAK6754	
55	represents human immune/hematopoietic and/or cancer-related	
56	sequences from the present invention. AAK4942 to AAK4950 and AAK4951	
57	represent sequences used in the exemplification of the present invention.	
58		
59	Sequence 10740 bp; 4259 A; 1965 G; 4229 C; 8007 T; 6 other:	
60		
61	Query Match	7.99; Score: 86; Dh: 22; Length: 10740
62	Best Local Similarity	51.99; Pval: 8.3e-7
63	Matches: 112; Conservative: 0; Mismatches: 100; Indels: 4; Gaps: 1	
64		
65	33 AGAAGAGTGAAGAGTACAGTATCTGGTCAAAATAGAGAGATTTTCTGCTGGCTTT	92
66	Db 7904 AGAAGATTTATTAATGATGACAAAGTTTCTGGAGCAAGATTTTCTAGTCTACCAAT	7845
67	93 TGAATACCTGACCGACAGACATGACAGTTCTAGAGACGCGCTTATCTGAAATATGGA	152
68	Db 7844 TAAAGAAATAGCAAAATGAAATTTTAAAGATGGAAGAGATCTCAAGACAGAGAGAGAG	7765
69	97 153 ATAAATCTGACGCTGATTTTGAAGAGATCTGATTTTCTGTTTCTGTTTCTGTTT	212
70	Db 7764 AGAAGATTAATGATGATGATTTCTGAGAAATTTTAAATATGATGATGATGATGAT	7724
71	97 213 GTCAG	249
72	Db 7724 TACTTAAAG	7699
73	RESULT 13	
74	AA161112/c	
75	Db AA161112 standard; cDNA: 817 bp.	
76	XX AC	
77	XX AA161112:	
78	01 07-DEC-2001 (first entry)	
79	XX XX	
80	Human breast cancer expressed polynucleotide 8569.	
81	XX XX	
82	Human breast cancer; cell marker; cytosolic; ss.	
83	XX XX	
84	Homo sapiens.	
85	XX XX	
86	W/200151628-A2.	
87	XX XX	
88	19-JUL-2001.	
89	XX XX	
90	10-JAN-2001; 2001WO-0500798.	
91	XX XX	
92	14-JAN-2000; 2000US-0176077.	
93	PR 14-MAR-2000; 2000US-0185167.	
94	PR 24-MAR-2000; 2000US-0192099.	
95	PR 23-MAR-2000; 2000US-0191480.	
96	PR 15-MAY-2000; 2000US-0205230.	
97	PR 09-JUN-2000; 2000US-0211315.	
98	PR 25-JUL-2000; 2000US-0220534.	
99	XX XX	
100	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
101	XX XX	
102	Lillie J. Xu Y, Wang Y, Steinmann K.	
103	XX XX	
104	WFI; 2001-451856/48.	
105	XX XX	
106	XX XX	
107	New peptide useful as a marker for the diagnosis of breast cancer.	
108	XX XX	
109	claim 1; Page 1550; 3695pp; English.	
110	XX XX	
111	XX XX	
112	XX XX	
113	XX XX	
114	XX XX	
115	XX XX	
116	XX XX	
117	XX XX	
118	XX XX	
119	XX XX	
120	XX XX	
121	XX XX	
122	XX XX	
123	XX XX	
124	XX XX	
125	XX XX	
126	XX XX	
127	XX XX	
128	XX XX	
129	XX XX	
130	XX XX	
131	XX XX	
132	XX XX	
133	XX XX	
134	XX XX	
135	XX XX	
136	XX XX	
137	XX XX	
138	XX XX	
139	XX XX	
140	XX XX	
141	XX XX	
142	XX XX	
143	XX XX	
144	XX XX	
145	XX XX	
146	XX XX	
147	XX XX	
148	XX XX	
149	XX XX	
150	XX XX	
151	XX XX	
152	XX XX	
153	XX XX	
154	XX XX	
155	XX XX	
156	XX XX	
157	XX XX	
158	XX XX	
159	XX XX	
160	XX XX	
161	XX XX	
162	XX XX	
163	XX XX	
164	XX XX	
165	XX XX	
166	XX XX	
167	XX XX	
168	XX XX	
169	XX XX	
170	XX XX	
171	XX XX	
172	XX XX	
173	XX XX	
174	XX XX	
175	XX XX	
176	XX XX	
177	XX XX	
178	XX XX	
179	XX XX	
180	XX XX	
181	XX XX	
182	XX XX	
183	XX XX	

XX
67 testing primers and probes for analysing diseases associated with
68 cytosine methylation state e.g. arthritis, cancer, aging,
69 arteriosclerosis comprising fragments of chemically modified genes
70 associated with cell cycle
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

Search completed: August 20, 2003, 14:45:39
Job time : 189.717 secs

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 81512)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS
TITLE
JOURNAL
3 (bases 1 to 81512)
DOE Joint Genome Institute.
Direct Submission
Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS
TITLE
JOURNAL
4 (bases 1 to 81512)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (01-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
5 (bases 1 to 81512)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (28-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 28, 2003 this sequence version replaced gi:20376952.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.3% of Sequence;
Estimated Total Number of Errors is 0.4.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 139.5kb). It is clipped at the overlaps with AC008378 and AC106731. The number of bases overlapped with AC008378 is 4442 bps and with AC106731 is 15413 bps.
Location/Qualifiers

FEATURES

source

1. 81512
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-281H14"
BASE COUNT 20011 a 19534 c 19980 g 21987 t
ORIGIN

Query Match 100.0%; Score 478; DB 9; Length 81512;
Best Local Similarity 100.0%; Pred. No. 3.8e-122; Indels 0; Gaps 0;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AGAGAAATCATATTACCCGATTCACAAAGAGCATAGAGATGTAACTGATCTTGT 60
|||||
46088 AGAGAAATCATATTACCCGATTCACAAAGAGCATAGAGATGTAACTGATCTTGT 46029
61 CAATAGGAGAGTTTTTTTCTTCCCTTTTGTAACTGACCCACAGAGCTGACAG 120
|||||
46028 CAATAGGAGAGTTTTTTTCTTCCCTTTTGTAACTGACCCACAGAGCTGACAG 45969
121 TTTCAGGAGAGCCCTTACCCGAAATAGGAAATTAATTCCTGGCACCCTTGATTGGCAAG 180
|||||
45968 TTTCAGGAGAGCCCTTACCCGAAATAGGAAATTAATTCCTGGCACCCTTGATTGGCAAG 45909
181 GGCAATGCTAATTTTTTTTCTTCTCCAGAGCTCTCAAAAAAATTTTTTTTAC 240
|||||
45908 GGCAATGCTAATTTTTTTTCTTCTCCAGAGCTCTCAAAAAAATTTTTTTTAC 45849
241 TAAAAACAGGATCCCGATGATGCTGATGTCCTCCCATTTAAACGTTAATTTTCAGGC 300
|||||
45848 TAAAAACAGGATCCCGATGATGCTGATGTCCTCCCATTTAAACGTTAATTTTCAGGC 45789
301 GTCCGCTACACTAATTTTCTTCAAACTGCATCGCAGCCGCTGGCCACAGATTCACCT 360
|||||
45788 GTCCGCTACACTAATTTTCAAACTGCATCGCAGCCGCTGGCCACAGATTCACCT 45729
361 AACAGCGCTCCAGAGACCTGCTCCGAGCTCTTTTCAGCGAGCATTTAATTTGAATCGG 420
|||||
45728 AACAGCGCTCCAGAGACCTGCTCCGAGCTCTTTTCAGCGAGCATTTAATTTGAATCGG 45669
|||||

Oy 421 ATGTGGCTCGTTTGGCCAGAGCTACCGGCTCGCGATAGGATCTCTCCACGACAC 478
|||||
Db 45668 ATGTGGCTCGTTTGGCCAGAGCTACCGGCTCGCGATAGGATCTCTCCACGACAC 45611

RESULT 2
AL669924/C
LOCUS
DEFINITION
Human DNA sequence from clone Xhdac-2L19 on chromosome 6, complete sequence.
ACCESSION
AL669924
VERSION
AL669924.9 GI:20068712
KEYWORDS
HTG.
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
1
Direct Submission
Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humqueresanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Apr 7, 2002 this sequence version replaced gi:19572797.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Xhdac-2L19 is from a CHO1-501 human bac - pgf cell line library VECTOR: pTARbac2.1

This sequence was generated from part of bacterial clone contigs constructed by the MHC Haplotype Consortium and collaborators. Further information can be found at
http://www.sanger.ac.uk/HGP/chr6/MHC.

FEATURES

source

1. 137345
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Xhdac-2L19"
/clone_11b="CHO1-501"
BASE COUNT 33377 a 33207 c 33325 g 37436 t
ORIGIN

Query Match 97.5%; Score 466; DB 9; Length 137345;
Best Local Similarity 99.8%; Pred. No. 8.7e-119;
Matches 477; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy 1 AGAGAAATCATTTACCCGATTCACAAAGAGCATAGAGATGTAACTGATCTTGT 60
|||||
Db 46314 AGAGAAATCATTTACCCGATTCACAAAGAGCATAGAGATGTAACTGATCTTGT 46255
61 CAATAGGAGAGTTTTTTTCTTCCCTTTTGTAACTGACCCACAGAGCTACAG 120
|||||
Db 46254 CAATAGGAGAGTTTTTTTCTTCCCTTTTGTAACTGACCCACAGAGCTACAG 46195
121 TTCTAGGAGAGCCCTTACCCGAAATAGGAAATTAATCTTGGCCACTGATTGGCAAG 180
|||||

Chemistry: Dye-Primer ET; 0% of reads
Chemistry: Dye-Terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 140428 bases at least Q40
Consensus quality: 146442 bases at least Q30
Consensus quality: 149661 bases at least Q20
Insert size: 217000; agarose-fp
Insert size: 153246; sum-of-contigs
Quality coverage: 1.93 in Q20 bases; agarose-fp
Quality coverage: 2.65 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1262: contig of 1262 bp in length
* 1263 1362: gap of unknown length
* 1363 2809: contig of 1447 bp in length
* 2810 2909: gap of unknown length
* 2910 4113: contig of 1204 bp in length
* 4114 4213: gap of unknown length
* 4214 5492: contig of 1279 bp in length
* 5493 5592: gap of unknown length
* 5593 6902: contig of 1310 bp in length
* 6903 7002: gap of unknown length
* 7003 8255: contig of 1253 bp in length
* 8256 8355: gap of unknown length
* 8356 9423: contig of 1068 bp in length
* 9424 9523: gap of unknown length
* 9524 10777: contig of 1254 bp in length
* 10778 10877: gap of unknown length
* 10878 12023: contig of 1146 bp in length
* 12024 12123: gap of unknown length
* 12124 13409: contig of 1286 bp in length
* 13410 13509: gap of unknown length
* 13510 15253: contig of 1744 bp in length
* 15254 15353: gap of unknown length
* 15354 16649: contig of 1296 bp in length
* 16650 16749: gap of unknown length
* 16750 18229: contig of 1480 bp in length
* 18230 18329: gap of unknown length
* 18330 19662: contig of 1333 bp in length
* 19663 19762: gap of unknown length
* 19763 21024: contig of 1262 bp in length
* 21025 21124: gap of unknown length
* 21125 22467: contig of 1343 bp in length
* 22468 22567: gap of unknown length
* 22569 23615: contig of 1048 bp in length
* 23616 23715: gap of unknown length
* 23716 25651: contig of 1936 bp in length
* 25652 25751: gap of unknown length
* 25752 27281: contig of 1530 bp in length
* 27282 27381: gap of unknown length
* 27382 29420: contig of 2039 bp in length
* 29421 29520: gap of unknown length
* 29521 30901: contig of 1381 bp in length
* 30902 31001: gap of unknown length
* 31002 32159: contig of 1158 bp in length
* 32160 32259: gap of unknown length
* 32260 33788: contig of 1529 bp in length
* 33789 33888: gap of unknown length
* 33889 35134: contig of 1246 bp in length
* 35135 35234: gap of unknown length
* 35235 36486: contig of 1252 bp in length
* 36487 36586: gap of unknown length
* 36587 38385: contig of 1799 bp in length
* 38386 38485: gap of unknown length
* 38486 39714: contig of 1229 bp in length
* 39715 39814: gap of unknown length

39815 41078: contig of 1264 bp in length
* 41079 41178: gap of unknown length
* 41179 42273: contig of 1095 bp in length
* 42274 42373: gap of unknown length
* 42374 44123: contig of 1750 bp in length
* 44124 44223: gap of unknown length
* 44224 45908: contig of 1685 bp in length
* 45909 46008: gap of unknown length
* 46009 47545: contig of 1537 bp in length
* 47546 47645: gap of unknown length
* 47646 50012: contig of 2367 bp in length
* 50013 50112: gap of unknown length
* 50113 51978: contig of 1866 bp in length
* 51979 52078: gap of unknown length
* 52079 54113: contig of 2035 bp in length
* 54114 54213: gap of unknown length
* 54214 55586: contig of 1373 bp in length
* 55587 55686: gap of unknown length
* 55687 57405: contig of 1719 bp in length
* 57406 57505: gap of unknown length
* 57506 58790: contig of 1285 bp in length
* 58791 58890: gap of unknown length
* 58891 60453: contig of 1563 bp in length
* 60454 60553: gap of unknown length
* 60554 62149: contig of 1536 bp in length
* 62150 62249: gap of unknown length
* 62250 63910: contig of 1661 bp in length
* 63911 64010: gap of unknown length
* 64011 65910: contig of 1900 bp in length
* 65911 65911: gap of unknown length
* 65912 66010: gap of unknown length
* 66011 67754: contig of 1744 bp in length
* 67755 67854: gap of unknown length
* 67855 69859: contig of 2005 bp in length
* 69860 69959: gap of unknown length
* 69960 71962: contig of 2003 bp in length
* 71963 72062: gap of unknown length
* 72063 74371: contig of 2309 bp in length
* 74372 74471: gap of unknown length
* 74472 76499: contig of 2028 bp in length
* 76500 76599: gap of unknown length
* 76600 78329: contig of 1730 bp in length
* 78330 78429: gap of unknown length
* 78430 81033: contig of 2604 bp in length
* 81034 81133: gap of unknown length
* 81134 82450: contig of 1317 bp in length
* 82451 82550: gap of unknown length
* 82551 85880: contig of 3330 bp in length
* 85881 85980: gap of unknown length
* 85981 89359: contig of 3979 bp in length
* 89360 90059: gap of unknown length
* 90060 92499: contig of 2440 bp in length
* 92500 92599: gap of unknown length
* 92600 95372: contig of 2773 bp in length
* 95373 95472: gap of unknown length
* 95473 98796: contig of 3324 bp in length
* 98797 98896: gap of unknown length
* 98897 101619: contig of 2733 bp in length
* 101620 101719: gap of unknown length
* 101720 105463: contig of 3744 bp in length
* 105464 105563: gap of unknown length
* 105564 108520: contig of 2957 bp in length
* 108521 108620: gap of unknown length
* 108621 111714: contig of 3094 bp in length
* 111715 111814: gap of unknown length
* 111815 115701: contig of 3887 bp in length
* 115702 115801: gap of unknown length
* 115802 117720: contig of 1919 bp in length
* 117721 117820: gap of unknown length
* 117821 121379: contig of 3459 bp in length
* 121380 121480: gap of unknown length
* 121481 125489: contig of 4110 bp in length
* 125490 125589: gap of unknown length
* 125590 128887: contig of 3308 bp in length

Db 20912 ATCCGAGGGCCAGCCGATTCATCTTAACAGACTCCAGAGACCTCCGCCAGAGCTGTTT 20971
QY 396 TCAGCGAGACATTTAATTGATCGATCGCTGCTTGGCCAGACGTCAACCGCTCGGCG 455
Db 20972 TCAGCAGACATTTAATTGATCGATCGCTGCTTGGCCAGACGTCAACCGCTCGGCG 21031
QY 456 ATAGGCATCTCTCCACGACAC 478
Db 21032 ATAGCATCTCTCCACGACAC 21054

RESULT 6
AF083133 8648 bp DNA linear ROD 08-DEC-2000
LOCUS AF083133
DEFINITION Mus musculus cardiac homeobox transcription factor (Kxx2-5) gene,
complete cds.
ACCESSION AF083133 GI:11602837
VERSION AF083133
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Tanaka,M., Wechsler,S.B., Lee,I.W., Yamasaki,N., Lawlits,J. and
Izumo,S.
TITLE Modular cis-acting elements regulate expression of the murine
Cxx/Kxx2-5 homeobox gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 8648)
AUTHORS Lee,I.W., Wechsler,S.B., Tanaka,M., Yamasaki,N. and Izumo,S.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1998) Cardiology, Beth Israel Deaconess Medical
Center, Harvard Medical School, 330 Brookline Avenue, Boston, MA
02215, USA

FEATURES
source 1..8648 Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129"
/db_xref="taxon:10090"
gene 1114..8114
/gene="Nkx2-5"
/note="synonym: Csx"
mRNA join(1114..1159,1885..1920,5171..5754,7132..8114)
/gene="Nkx2-5"
/product="cardiac homeobox transcription factor"
/note="alternatively spliced"
mRNA join(1828..1920,5171..5754,7132..8114)
/gene="Nkx2-5"
/product="cardiac homeobox transcription factor"
/note="alternatively spliced"
mRNA join(5100..5754,7132..8114)
/gene="Nkx2-5"
/product="cardiac homeobox transcription factor"
/note="alternatively spliced"
CDS join(5424..5754,7132..7757)
/gene="Nkx2-5"
/function="involved in heart development"
/note="timman homolog; Nkx family"
/codon_start=1
/product="cardiac homeobox transcription factor"
/protein_id="AAG38875.1"
/db_xref="GI:11602838"
/translation="MPPSPALPTPFPSVKYKILNLEOQRLASGDLASRLATLAPAS
CMIAAFREAYSGEMASASGLAEIRAEKGRAPRSPKSPAPPAATFPFGAGDDPA
KDPRAKKEICAIQKAVELDKAETDGARPRARRKPRVLESAQVLELEKQQR
YLSPRKQILASVILKISTQYKIMFQNRKRCRQRQDTELLGPPPPARIAVY
LVKDGKICIDPAAYPAAYGVLNAYGNAYPSYSGAACSPPGSCAAYAAPPAAH
APASANSNPNFVGVDINTVQSPMPGNGSVSLHGIRAM"

BASE COUNT 2041 a 2336 c 2179 g 2091 t 1 others

Query Match 38.1%; Score 182.2; DB 10; Length 8648;
Best Local Similarity 71.6%; Pred. No. 9,6e-40;
Matches 317; Conservative 0; Mismatches 103; Indels 23; Gaps 5;

QY 36 GAGCTAACAGTACTGATCTTGTTCAAAATAGGAGAGTTTTCCTTCCTTTTGT 95
Db 3009 GGGTGGGAAGACACATGATTTGTTCAAAATAGGAGAGTTTTCCTTCCTTTTGT 3063
QY 96 AACACCTGACCCAGAGAGTACAGTTCAGAGAGCCGCCCTTACCGAAATAGGAATA 155
Db 3064 GACACCTAGCCACAGATTTTGCAACTCT-6GAAGCCCTATATTCGAAAA-AAATGATA 3121
QY 156 AATCCCTGCACCTTGATTTGCAAGGCAATCTTAATTTTCTTCTTCAGAGCTCTC 215
Db 3122 AATCCTCAACATGATTTGCAAGGCAATCTTAATTTTCTTCTTCAGAGCTCTC 3171
QY 216 AAAAAAAAAAAAAAAAACTTACTTAAACAGGATCCCGATGTAGCTTCGATGCC 275
Db 3172 TTTTAAAAAGACTTAAACAGGATCCCGATGTAGGAGGCGCCCAATATGCTC-----CC 3225
QY 276 CCATTTAAAGGTATATTTTCAGGCGTCCGCTCACACTAATCTTCAAACTGCATGCCG 335
Db 3226 CCATTTAAAGGTATATTTTCAGGCGTCCGCTCACACTAATCTTCAAACTGCATGCCG 3285
QY 336 AGCCGCTGCGCAGCAGATTCCTTACAGGCGCTCCAGAGACCTTCGTCGAGCTCTT 395
Db 3286 ATCCGAGGGGAGCGAGATTCCTTACAGGAGCTCCAGAGACCTTCGTCGAGCTCTT 3345
QY 396 TCAGCGAGACATTTAATTGAATCGATCGCTGCTTGGCCAGAGCTCGCCGTCGCGG 455
Db 3346 TCAGCAGACATTTAATTGAATCGATCGCTGCTTGGCCAGAGCTCGCCGTCGCGG 3405

QY 456 ATAGCATCTCTCCACGACAC 478
Db 3406 ATAGCATCTCTCCACGACAC 3428

RESULT 7
AC119699/c 171110 bp DNA linear HTG 13-NOV-2002
LOCUS AC119699/c
DEFINITION Rattus norvegicus clone CH230-466D21, WORKING DRAFT SEQUENCE, 5
unoriented pieces.
ACCESSION AC119699 GI:24941740
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 171110)
Muzny,D.,Meire., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barstead,M., Benahmed,F.,
Bisvalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Bunay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinu,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrergergs,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunther,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huylk,S., Hume,J., Idelbird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

LOCUS	AC102990/c	241724	bp	DNA	linear	HTG 10-MAY-2003	
DEFINITION	Rattus norvegicus clone CH230-53024,			WORKING DRAFT		SEQUENCE.	
ACCESSION	AC102990						
VERSION	AC102990.6	GI:30521867					
KEYWORDS	HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.						
SOURCE	Rattus norvegicus (Norway rat)						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.						
REFERENCE	1 (bases 1 to 241724)						
AUTHORS	Muzny,D.,Marie., Metker,M.,Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D., Ayvaldebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benhammed,F., Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Butrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Galla,R., Garcia,A., Garner,T., Garza,M., Gevorgian,E., Geer,K., Gitter,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlik,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Hultyk,S., Hume,U., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Kapachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kovis,C., Krieff,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinec,E., Mathlney,S., McLeod,M.P., McNeill,T.B., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundayase,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwokilemeh,O., Okwou,G., Olarpunsgoon,A., Pal,S., Parks,K., Paternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polndexter,A., Popovic,D., Prims,E., Pu,L.,-L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajd,S., Smed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,U., Steimle,T., Strong,R., Sutton,A., Svalik,A., Taboi,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakut,S., Yen,O., Yoon,L., Yoon,Y., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.						
TITLE	Direct Submission						
REFERENCE	2 (bases 1 to 241724)						
AUTHORS	Worley,K.C.						
TITLE	Direct Submission						
JOURNAL	Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA						
REFERENCE	3 (bases 1 to 241724)						
AUTHORS	Rat Genome Sequencing Consortium.						
TITLE	Direct Submission						
JOURNAL	Submitted (10-MAY-2003) Human Genome Sequencing Center, Department						

```

COMMENT
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced g1:22855562.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHEJ
Center clone name: CH230-53024
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 233518 bases at least Q40
Consensus quality: 235377 bases at least Q30
Consensus quality: 236878 bases at least Q20
Estimated insert size: 245960; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as runs of N, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
  1 241724: contig of 241724 bp in length.
Location/Qualifiers
  1. 241724
  /organism="Rattus norvegicus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10116"
  /clone="CH230-53024"
  1. 2211
misc_feature
  /note="wgs_contig"
BASE COUNT  60533 a 53609 c 56345 g 67996 t 3241 others
ORIGIN
Query Match      38.0%; Score 181.8; DB 2; Length 241724;
Best Local Similarity 68.6%; Pred. No. 1.5e-39;
Matches 319; Conservative 0; Mismatches 122; Indels 24; Gaps 4;

  15 CCGATTCCAAAGACATAGAGAGTGTAACAGTCACTGATCTTTGTTCAATAGGGAGAT 74
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168674 CCTTTTCACCAAGGTGTAGGGCGCGGAAAGTCACTGATTTGTTCAATTAAGAAAGT 168615

  75 TTTTTCCTCCCTTTTGTATACACGTACCCACAGAGCTACACTTCTAGGAAGCCCC 134
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168614 TTT-----CTTCCCTTTTCTTATACACTTACCCACAGATCTGTCAACTCT -GGAGCCTT 168561

  135 CTTACCCGAAATATAGAAATAAATCTTGCCACTTGATTTGTCAAAGGCAATGCTAATT 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168560 TATATCCGGGAAAGGTAGTAATATCCCCACCATATTGATTGTCAAAGGAATGTCCTCC 168501

  195 TTTTCTTCTCGAGGCTCTCAAAAAAAAAAAAAAAAAAACTTACTTAAACAGGATC 254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168500 TCAGACATCATTTATTTAAAAAAGCGTAAACAGGAGATCCCGATGTGCCCC 168441

```

[illegible]

COMMENT

11116
JOURNAL

Phunkiang, P., Pierre, N., Raymond, C., Fecteau, R., Fiset, J., Goff, P., Roman, J., Roy, A., Schaefer, S., Schepack, R., Senneker, S., Sperry, P., Smith, C., Spencer, B., Stader, J., Steinhilber, N., Staudacher, N., Thomas, J., Teysseie, S., Theodorou, J., Totham, K., Travers, M., Vassiliadis, B., Viel, R., Vo, A., Wilson, B., Wu, X., Wurmhuber, Y., Yoneda, S., Zaitsev, D., Zembek, L., Zimmer, A. and Zou, J. M.

Direct Submission

Submitted (08-Jun-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 9, 2002 this sequence version replaced 92121700715.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/MW/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WMK
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 531_C_8
Center phone name: 531_C_8

Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: dye-termination; Big Dye3; 100% of reads
Assembly program: Phrap; version: 6.90271
Consensus quality: 165910 bases at least 40
Consensus quality: 16597 bases at least 20
Consensus quality: 161126 bases at least 92
Insert size: 17000-40000 bp
Insert size: 161647; sum of contigs
Quality coverage: 4.7 in gzo bases; adarose*
Quality coverage: 5.0 in gzo bases; sum of contigs

* NOTE: This is a "working draft" sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

*	1	1212; contig of 1212 bp in length
*	1213	1312; gap of 100 bp
*	1413	2490; contig of 1478 bp in length
*	2991	3090; gap of 100 bp
*	3091	5404; contig of 2414 bp in length
*	5405	5504; gap of 100 bp
*	5505	8570; contig of 3066 bp in length
*	8571	8670; gap of 100 bp
*	8671	12334; contig of 3664 bp in length
*	12345	12434; gap of 100 bp
*	12435	17442; contig of 4908 bp in length
*	17443	17442; gap of 100 bp
*	17443	21556; contig of 4714 bp in length
*	21157	21556; gap of 100 bp
*	21557	25915; contig of 4459 bp in length
*	25916	26016; gap of 100 bp
*	26016	32932; contig of 6917 bp in length
*	32933	33032; gap of 100 bp
*	33033	40214; contig of 7162 bp in length
*	40215	40914; gap of 100 bp
*	40315	50486; contig of 10072 bp in length
*	50387	50486; gap of 100 bp
*	50487	65419; contig of 14884 bp in length
*	65420	65419; gap of 100 bp
*	65420	72608; contig of 13989 bp in length
*	72609	73508; gap of 100 bp
*	73509	98219; contig of 16671 bp in length
*	98219	98219; gap of 100 bp
*	98220	114784; contig of 17794 bp in length
*	114784	114784; gap of 100 bp
*	114784	164337; contig of 14079 bp in length

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
-----Project Information-----
Center project name: L6875
Center clone name: 404_F_23

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allow
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 783: contig of 783 bp in length
* 784 883: gap of 100 bp
* 884 1715: contig of 832 bp in length
* 1716 1815: gap of 100 bp
* 1816 2635: contig of 820 bp in length
* 2636 2735: gap of 100 bp
* 2736 3523: contig of 788 bp in length
* 3524 3623: gap of 100 bp
* 3624 4437: contig of 814 bp in length
* 4438 4537: gap of 100 bp
* 4538 5363: contig of 826 bp in length
* 5364 5463: gap of 100 bp
* 5464 6265: contig of 802 bp in length
* 6266 6365: gap of 100 bp
* 6366 7189: contig of 824 bp in length
* 7190 7289: gap of 100 bp
* 7290 8104: contig of 815 bp in length
* 8105 8204: gap of 100 bp
* 8205 9008: contig of 804 bp in length
* 9009 9108: gap of 100 bp
* 9109 9939: contig of 831 bp in length
* 9940 10039: gap of 100 bp
* 10040 10873: contig of 834 bp in length
* 10874 10973: gap of 100 bp
* 10974 11801: contig of 828 bp in length
* 11802 11901: gap of 100 bp
* 11902 12697: contig of 796 bp in length
* 12698 12797: gap of 100 bp
* 12798 13591: contig of 794 bp in length
* 13592 13691: gap of 100 bp
* 13692 14505: contig of 814 bp in length
* 14506 14605: gap of 100 bp
* 14606 15432: contig of 827 bp in length
* 15433 15532: gap of 100 bp
* 15533 16360: contig of 828 bp in length
* 16361 16460: gap of 100 bp
* 16461 17278: contig of 818 bp in length
* 17279 17378: gap of 100 bp
* 17379 18214: contig of 836 bp in length
* 18215 18314: gap of 100 bp
* 18315 19152: contig of 838 bp in length
* 19153 19252: gap of 100 bp
* 19253 20088: contig of 836 bp in length
* 20089 20188: gap of 100 bp
* 20189 20963: contig of 775 bp in length
* 20964 21063: gap of 100 bp
* 21064 21894: contig of 831 bp in length
* 21895 21994: gap of 100 bp
* 21995 22759: contig of 765 bp in length
* 22760 22859: gap of 100 bp
* 22860 23645: contig of 786 bp in length
* 23646 23745: gap of 100 bp
* 23746 24541: contig of 796 bp in length
* 24542 24641: gap of 100 bp

24642 25447: contig of 806 bp in length
* 25448 25547: gap of 100 bp
* 25548 26373: contig of 826 bp in length
* 26374 26473: gap of 100 bp
* 26474 27287: contig of 814 bp in length
* 27288 27387: gap of 100 bp
* 27388 28210: contig of 823 bp in length
* 28211 28310: gap of 100 bp
* 28311 29107: contig of 797 bp in length
* 29108 29207: gap of 100 bp
* 29208 30038: contig of 831 bp in length
* 30039 30138: gap of 100 bp
* 30139 30966: contig of 828 bp in length
* 30967 31066: gap of 100 bp
* 31067 31872: contig of 806 bp in length
* 31873 31972: gap of 100 bp
* 31973 32772: contig of 800 bp in length
* 32773 32872: gap of 100 bp
* 32873 33672: contig of 800 bp in length
* 33673 33772: gap of 100 bp
* 33773 34587: contig of 815 bp in length
* 34588 34687: gap of 100 bp
* 34688 35506: contig of 819 bp in length
* 35507 35606: gap of 100 bp
* 35607 36418: contig of 812 bp in length
* 36419 36518: gap of 100 bp
* 36519 37340: contig of 822 bp in length
* 37341 37440: gap of 100 bp
* 37441 38266: contig of 826 bp in length
* 38267 38366: gap of 100 bp
* 38367 39191: contig of 825 bp in length
* 39192 39291: gap of 100 bp
* 39292 40075: contig of 784 bp in length
* 40076 40175: gap of 100 bp
* 40176 41005: contig of 830 bp in length
* 41006 41105: gap of 100 bp
* 41106 41919: contig of 814 bp in length
* 41920 42020: gap of 100 bp
* 42020 42822: contig of 803 bp in length
* 42823 42922: gap of 100 bp
* 42923 43745: contig of 823 bp in length
* 43746 43845: gap of 100 bp
* 43846 44653: contig of 808 bp in length
* 44654 44753: gap of 100 bp
* 44754 45563: contig of 810 bp in length
* 45564 45663: gap of 100 bp
* 45664 46491: contig of 828 bp in length
* 46492 46591: gap of 100 bp
* 46592 47432: contig of 841 bp in length
* 47433 47532: gap of 100 bp
* 47533 48374: contig of 842 bp in length
* 48375 48474: gap of 100 bp
* 48475 49273: contig of 799 bp in length
* 49274 49373: gap of 100 bp
* 49374 50190: contig of 817 bp in length
* 50191 50290: gap of 100 bp
* 50291 51083: contig of 793 bp in length
* 51084 51183: gap of 100 bp
* 51184 51939: contig of 756 bp in length
* 51940 52039: gap of 100 bp
* 52040 52848: contig of 809 bp in length
* 52849 52948: gap of 100 bp
* 52949 53763: contig of 815 bp in length
* 53764 53863: gap of 100 bp
* 53864 54698: contig of 835 bp in length
* 54699 54798: gap of 100 bp
* 54799 55573: contig of 775 bp in length
* 55574 55673: gap of 100 bp
* 55674 56508: contig of 835 bp in length
* 56509 56608: gap of 100 bp
* 56609 57439: contig of 831 bp in length
* 57440 57539: gap of 100 bp
* 57540 58355: contig of 816 bp in length

AC116644	AC116644	65419 bp	DNA	ligent	PR1 27-MAY-2002
DEFINITION	Homo sapiens BAC clone RP11-462013 from 4, complete sequence.				
AC116644	AC116644.5	61,204,296 bp			
VERSION	H13.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Mammalian Molecular Cloning Library (GenBank: W100000) F1000000000				
AUTHORS	1. (bases 1 to 65419)				
ADDITIONAL	Submitted (to MAR-2002) Genome Sequencing Center, Washington				
ADDITIONAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
ADDITIONAL	MO 63104, USA				
ADDITIONAL	4 (bases 1 to 65419)				
ADDITIONAL	Submitted (to APR-2002) Genome Sequencing Center, Washington				
ADDITIONAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
ADDITIONAL	MO 63104, USA				
ADDITIONAL	5 (bases 1 to 65419)				
ADDITIONAL	Submitted (to MAY-2002) Genome Sequencing Center, Washington				
ADDITIONAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
ADDITIONAL	MO 63104, USA				
ADDITIONAL	6 (bases 1 to 65419)				

AC116644	AC116644	65419 bp	DNA	ligent	PR1 27-MAY-2002
DEFINITION	Homo sapiens BAC clone RP11-462013 from 4, complete sequence.				
AC116644	AC116644.5	61,204,296 bp			
VERSION	H13.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Mammalian Molecular Cloning Library (GenBank: W100000) F1000000000				
AUTHORS	1. (bases 1 to 65419)				
ADDITIONAL	Submitted (to MAR-2002) Genome Sequencing Center, Washington				
ADDITIONAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
ADDITIONAL	MO 63104, USA				
ADDITIONAL	4 (bases 1 to 65419)				
ADDITIONAL	Submitted (to APR-2002) Genome Sequencing Center, Washington				
ADDITIONAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
ADDITIONAL	MO 63104, USA				
ADDITIONAL	5 (bases 1 to 65419)				
ADDITIONAL	Submitted (to MAY-2002) Genome Sequencing Center, Washington				
ADDITIONAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
ADDITIONAL	MO 63104, USA				
ADDITIONAL	6 (bases 1 to 65419)				

repeat_region	5587..6005	/rpt_family="ERV1"
repeat_region	6006..6518	/rpt_family="MaLR"
repeat_region	6519..7199	/rpt_family="ERV1"
repeat_region	7200..7494	/rpt_family="Alu"
repeat_region	7495..7987	/rpt_family="ERV1"
repeat_region	8014..8807	/rpt_family="ERV1"
repeat_region	8857..8948	/rpt_family="MaLR"
repeat_region	9002..9396	/rpt_family="ERV1"
repeat_region	10045..10493	/rpt_family="ERV1"
repeat_region	11114..11591	/rpt_family="MaLR"
repeat_region	12897..13054	/rpt_family="L1"
repeat_region	13078..13108	/rpt_family="AT-rich"
repeat_region	13218..13321	/rpt_family="MaLR"
repeat_region	13322..13662	/rpt_family="MaLR"
repeat_region	15233..15565	/rpt_family="MaLR"
repeat_region	15366..15739	/rpt_family="MaLR"
repeat_region	16034..16063	/rpt_family="AT-rich"
repeat_region	16114..16158	/rpt_family="TIG"
repeat_region	16365..16827	/rpt_family="ERV1"
repeat_region	16885..16922	/rpt_family="CA"
repeat_region	21089..21516	/rpt_family="ERV1"
repeat_region	21517..21812	/rpt_family="Alu"
repeat_region	21813..22058	/rpt_family="ERV1"
repeat_region	22059..22444	/rpt_family="L1"
repeat_region	22425..23566	/rpt_family="MER2_type"
repeat_region	23557..24993	/rpt_family="L1"
repeat_region	25147..25565	/rpt_family="L1"
repeat_region	25584..26360	/rpt_family="L1"
repeat_region	26367..27217	/rpt_family="L1"
repeat_region	27255..27587	/rpt_family="L1"
repeat_region	27588..27866	/rpt_family="Alu"
repeat_region	27870..27891	/rpt_family="AAAA"
repeat_region	27892..28015	/rpt_family="L1"
repeat_region	28758..29098	/rpt_family="Alu"
repeat_region	30348..30656	/rpt_family="Alu"
repeat_region	30709..30799	/rpt_family="AT-rich"
repeat_region	31174..31384	

```
repeat_region
/rpt_family="A1u"
31697. .31896
/rpt_family="CR1"
repeat_region
31904. .31951
/rpt_family="ERV1"
repeat_region
32005. .32473
/rpt_family="L1"
repeat_region
32598. .32707
/rpt_family="L1"
repeat_region
32792. .32874
/rpt_family="(TA)n"
repeat_region
32992. .33026
/rpt_family="AT_rich"
repeat_region
33571. .33598
/rpt_family="AT_rich"
repeat_region
33665. .33708
```

Query Match	8.7%	Score 41.8	DB 9	Length 85439
Best Local Similarity	51.3%	Pred. No. 1.4		
Matches	97	Conservative	0	Mismatches 92
				Indels 0
				Gaps 0
QY	57	TGTCACATATAGGAGAGTTT	TTTTTCTCCCTTTTGTAA	CACCTGACCCAGAGACTG 116
Db	1454	TTTCTAAACAAAAACATATT	TGCTGGTCCCTTAACATG	TCATCACAACAGGTCCTG 1513
QY	117	ACAAGTTTAGAGAACCCCTT	ACCCGAAATATAGAAATAA	TCCCTGGCACCCTGATTTG 176
Db	1514	ACAGAGACAATATATCCACC	CTCTTAATATATGTGAAAA	AAATTTATGAACTCTAATTGAA 1573
QY	177	CAAGGCAATGCTAATTTT	TTCTTCCAGAGCTCTCAAA	AAAAAAAAAAAAAAAAAAACC 236
Db	1574	AGAGGGGAAACTAACAATG	TTATATACAAAAGCTCCAA	CAATATTTAAAGTGAAC 1633
QY	237	TTACTAAAA 245		
Db	1634	TTAGTCAAA 1642		

Search completed: August 20, 2003, 12:34:45
Job time : 2469.3 secs


```
source
1..577
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HFWAHE01"
/lissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="HTF"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      165 a      152 c      157 g      103 t
ORIGIN

Query Match      24.1%; Score 258; DB 9; Length 577;
Best Local Similarity 87.6%; Pred. No. 6.7e+02;
Matches 282; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 626 GGCCGGGGCCATGCTCAGCCCTGTATCCAGCACTTTAGAGGCCGAGCGGCCGA 685
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70 GGCCGGGGCGGTGGCTCAGCCCTGTATCCAGCACTTTGGAGGCCGAGCGGCCAGA 129
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 686 TCACCTGAGTGGGAGTTGCGACACAGCCCTGGCCACATGCTGTAACCCCTGTCTCTACA 745
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 130 TCACCTGAGTGGGAGTTGCCAGACAGCCCTGACCAACATGAGAGAACCCCTGTCTCTACT 189
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 746 AAAAAATACAAAATTAGCCGGGCGCATGGTGATCGCTGTATCCAGCACTTACTCTGGGAG 805
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 190 AAAAAATACAAAATTAGCCGGGCGCATGGTGATCGCTGTATCCAGCACTTACTCTGGTAG 249
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 806 GCTGAGCGAGAGAAATCGCTTGAACCCGGGAGCGGAGGTTGCACTGAGCCGAGATCACA 865
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 250 GCGGAGGCGAGAGAAATCGCTTGAACCCGGGAGCGGAGGTTGCACTGAGCCGAGATCAGG 309
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 866 CCACCTGACCTCCAGCTTGGGCGACAGAGCAAAATTCGCTAAAAAATATAAATAAT 925
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 310 CCACCTGACCTCCAGCTTGGGCGACAGAGCAAAATTCGCTCAAAAAAAGAAAAA 369
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 926 AAAATGATATTAATTAAGCCCATCA 947
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 370 AGAAAAAGAAATTAAGCCAGGCA 391
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 2
AV730440      661 bp      mRNA      linear      EST 17-OCT-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 661)
Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
Li,N., Qian,B., Liu,F., Ou,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA HTF clones
Unpublished
TITLE
JOURNAL
COMMENT
Contact: Zenguan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..661
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
/clone="HFWAHE01"
/lissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="HTF"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      183 a      178 c      179 g      116 t      5 others
ORIGIN

Query Match      24.1%; Score 258; DB 9; Length 661;
Best Local Similarity 87.6%; Pred. No. 6e+02;
Matches 282; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 626 GGCCGGGGCCATGCTCAGCCCTGTATCCAGCACTTTAGAGGCCGAGCGGCCGA 685
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 70 GGCCGGGGCGGTGGCTCAGCCCTGTATCCAGCACTTTGGAGGCCGAGCGGCCAGA 129
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 686 TCACCTGAGTGGGAGTTGCGACACAGCCCTGGCCACATGCTGTAACCCCTGTCTCTACA 745
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 130 TCACCTGAGTGGGAGTTGCCAGACAGCCCTGACCAACATGAGAGAACCCCTGTCTCTACT 189
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 746 AAAAAATACAAAATTAGCCGGGCGCATGGTGATCGCTGTATCCAGCACTTACTCTGGGAG 805
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 190 AAAAAATACAAAATTAGCCGGGCGCATGGTGATCGCTGTATCCAGCACTTACTCTGGTAG 249
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 806 GCTGAGCGAGAGAAATCGCTTGAACCCGGGAGCGGAGGTTGCACTGAGCCGAGATCACA 865
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 250 GCGGAGGCGAGAGAAATCGCTTGAACCCGGGAGCGGAGGTTGCACTGAGCCGAGATCAGG 309
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 866 CCACCTGACCTCCAGCTTGGGCGACAGAGCAAAATTCGCTAAAAAATATAAATAAT 925
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 310 CCACCTGACCTCCAGCTTGGGCGACAGAGCAAAATTCGCTCAAAAAAAGAAAAA 369
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 926 AAAATGATATTAATTAAGCCCATCA 947
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 370 AGAAAAAGAAATTAAGCCAGGCA 391
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 3
AM469140/c      437 bp      mRNA      linear      EST 24-FEB-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 437)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from gibco
High quality sequence stop: 420.
Location/Qualifiers
1..437
/organism="Homo sapiens"
```

```

/Model_Type "GENA"
/Ad_Xref "taxref:756"
/OT_ID "IMAG:200375"
/Tissue_Type "poorly differentiated adenocarcinoma with
slight foci of features"
/Gen_host "HHLH"
/Clone_ID "NCI_GARP_Gas4"
/Host_Organ "Stomach; Vector: PCMV-Spork16; Site:1; Salt:1;
Site:2; Salt:1; stored undifferentially. Primary: oligo of
Average Insert: size 1.59 kb. Life Technologies catalog #:
11533-014"

```

Query Match	2/44	Score 250.6	DB 9	Length 447
Post Local Similarity	8/48	Prod. No. 1.0e+03		
Matches 285	0	Mismatches 57	Indels 0	Gaps 0

[illegible]

684 GATGACCTTAAATTCAAGCCACTGGACAATGGGAAGCGGTCTCTA 74
CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT
285 CATTACTTAAGTCGTAATTAAGACCAAGCTTGTAAGCTGGTAGAACCTTGCTTTA 227

744 CAAAATACAAAAATTAAAGGACATCTGATTCCTGTCTTGATCCCAACTATCTGGG 804
+ |||||TTTTTTT|||||TTTTTTTTTTT|||TTTTTTTTTTTTTTTTTTTTTTTTTTT
726 TAAATAAATCAAAATTAAGCATCTGATCTCTTCTGATCTGATTCGAATCTATCGCG 167

804 AGCTTATCTACAGCAATCTTTTACGGCGGACATCGAGGCTTCAGTATGATGAGATCA 644

106 AGGAGAAATCTTCAAAAATTGTTCACAGCGAAGGGACAGCCTGCAGCGTTTG
864 CACTCATTAATCCCAAGCTTGGGATAAACAGGCCAATATTCCTTTAAAAAAATAAAAATAA 924

[illegible]

45. AAAAAAAAAAAGAGGAGTAATAATGCGAGTTTGGAGT 5

[illegible][illegible]

Homo sapiens (Human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE
ARTICLES
TITLE
National Institute of Health. Mammalian Gene Collection (MGC).
1 (bases 1 to 789)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 789)

PERSONAL COMMENT: Dr. Michael J. Fox
Contact: Fox, Michael J., Ph.D.
Email: fox@fox.rockefeller.edu
Phone: 212-327-7000 ext. 200

CLONETECH Laboratories, Inc.,
cDNA Library Preparation; CLONETECH Laboratories, Inc.,
cDNA Library Arrayed by the J.M.A.G.E. Consortium (J.M.N.)
DNA Sequenced by: Ingrid O. Odominis, Inc.

“Joint distribution M^* ” and distribution information can be found through the TMA.I.T. website (<http://imae.berkeley.edu>)

H₁₆h 141.12 deg 100.54 arc 54.1.
 1.033' 1.034/1.011 arc 1.
 1.038

```

/mol_type "rRNA"
/db_xref "taxon:9606"
/taxon "MARCH_407470"
/molecule_type "ribosomal_rRNA"
/lab_host "HISB (1) phage T-818ant"
/version "1.0-NIH_M06_60"
/name "Orphan; prostate; Vector: pMR-LIB (Clontech);
Site: S11 (aqueous/cyano); Site: z-S11 (aqueous/NaF)";
Double-stranded cDNA was prepared from total RNA,
5' and 3' adaptors were used in combination as follows: 5'
adaptor sequence: 5'-CAAGCATTATGACCTTGGTAACT
(sequence: 5' ATTTAATGATTAAGATTTGAATTCCTGAT
(Where H = A, C, G and N = A, C, G, T)). A 20bp
insert size (1.5 kb; range 0.4-3 kb), 147 bp genes
contained inserts by PCR. This library was screened for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: This is a NIH-MD
library".

```

Wong Match	23-48	Score 250-8	26-10	Leung 3-0
------------	-------	-------------	-------	-----------

610 AATTAAACCATGAGTGGGTGGATTCCTCAATCGTAATCCATCATTAGGA 569
Post Local Similarity: 85.6%; Prod. No. 7-20-27.
Matches: 279; Conservative: 0; Mismatches: 47; Indels: 0; Gaps: 0.

177 AAAAAAAAAAGGCTTAAATTAATAAATTGGTTTGTGAATTTCTAATTTATATATATTTTAAAA 236

[illegible][illegible]

790 CCGACGATTCGGAGGCTTGAAGCAGTAATAATTCTAAAGCTGTATGTTAATTTCTA 849
||||| + ||||| + ||||| + ||||| + ||||| + ||||| + ||||| + ||||| + ||||| + |||||
18 CCAATTAATTCGAAATCTTGAGGCTAATAAATCTCTAAAGCTATAGTCTAATTTCTA 476

[illegible]

```

910) AAAAAIAAAIAAAIAAAIAA 5+5
      | | | | | | | | | | | |
177) AAAAAAAAAAAAAAAAAAAAAA 5+7

```

Accession	Gene	Size (bp)	Accession	Gene	Size (bp)
AF208262	ADAM28	600	AF208262	ADAM28	600
AF208263	ADAM29	600	AF208263	ADAM29	600
AF208264	ADAM30	600	AF208264	ADAM30	600
AF208265	ADAM31	600	AF208265	ADAM31	600
AF208266	ADAM32	600	AF208266	ADAM32	600
AF208267	ADAM33	600	AF208267	ADAM33	600
AF208268	ADAM34	600	AF208268	ADAM34	600
AF208269	ADAM35	600	AF208269	ADAM35	600
AF208270	ADAM36	600	AF208270	ADAM36	600
AF208271	ADAM37	600	AF208271	ADAM37	600
AF208272	ADAM38	600	AF208272	ADAM38	600
AF208273	ADAM39	600	AF208273	ADAM39	600
AF208274	ADAM40	600	AF208274	ADAM40	600
AF208275	ADAM41	600	AF208275	ADAM41	600
AF208276	ADAM42	600	AF208276	ADAM42	600
AF208277	ADAM43	600	AF208277	ADAM43	600
AF208278	ADAM44	600	AF208278	ADAM44	600
AF208279	ADAM45	600	AF208279	ADAM45	600
AF208280	ADAM46	600	AF208280	ADAM46	600
AF208281	ADAM47	600	AF208281	ADAM47	600
AF208282	ADAM48	600	AF208282	ADAM48	600
AF208283	ADAM49	600	AF208283	ADAM49	600
AF208284	ADAM50	600	AF208284	ADAM50	600
AF208285	ADAM51	600	AF208285	ADAM51	600
AF208286	ADAM52	600	AF208286	ADAM52	600
AF208287	ADAM53	600	AF208287	ADAM53	600
AF208288	ADAM54	600	AF208288	ADAM54	600
AF208289	ADAM55	600	AF208289	ADAM55	600
AF208290	ADAM56	600	AF208290	ADAM56	600
AF208291	ADAM57	600	AF208291	ADAM57	600
AF208292	ADAM58	600	AF208292	ADAM58	600
AF208293	ADAM59	600	AF208293	ADAM59	600
AF208294	ADAM60	600	AF208294	ADAM60	600
AF208295	ADAM61	600	AF208295	ADAM61	600
AF208296	ADAM62	600	AF208296	ADAM62	600
AF208297	ADAM63	600	AF208297	ADAM63	600
AF208298	ADAM64	600	AF208298	ADAM64	600
AF208299	ADAM65	600	AF208299	ADAM65	600
AF208300	ADAM66	600	AF208300	ADAM66	600
AF208301	ADAM67	600	AF208301	ADAM67	600
AF208302	ADAM68	600	AF208302	ADAM68	600
AF208303	ADAM69	600	AF208303	ADAM69	600
AF208304	ADAM70	600	AF208304	ADAM70	600
AF208305	ADAM71	600	AF208305	ADAM71	600
AF208306	ADAM72	600	AF208306	ADAM72	600
AF208307	ADAM73	600	AF208307	ADAM73	600
AF208308	ADAM74	600	AF208308	ADAM74	600
AF208309	ADAM75	600	AF208309	ADAM75	600
AF208310	ADAM76	600	AF208310	ADAM76	600
AF208311	ADAM77	600	AF208311	ADAM77	600
AF208312	ADAM78	600	AF208312	ADAM78	600
AF208313	ADAM79	600	AF208313	ADAM79	600
AF208314	ADAM80	600	AF208314	ADAM80	600
AF208315	ADAM81	600	AF208315	ADAM81	600
AF208316	ADAM82	600	AF208316	ADAM82	600
AF208317	ADAM83	600	AF208317	ADAM83	600
AF208318	ADAM84	600	AF208318		

ATLANTA	AL28282
VERMONT	AL28282.1
NEW YORK	GL: 4670124
	1255.

[illegible]

REFERENCE	ADDITIONAL
1 (bases 1 to 500)	
Adams, M. D., Kousloy, S. D., Zhao, S., Russ, S., Liber, K., Madden, K., Berry, K., Chandler, R., Salas, E., Wilder, J., de Jongh, P., and Vetter, J. C.	
Use of human EAV and SV40 vectors for poliovirus RNA replication	

Unpublished
Contact: Mark Adams
Department of Education for Services
to Children, Youth and Families

^{13}C NMR (CDCl₃) δ : 101.6 (d, C-1), 100.6 (d, C-2), 100.4 (d, C-3), 100.3 (d, C-4), 100.2 (d, C-5), 100.1 (d, C-6), 100.0 (d, C-7), 99.9 (d, C-8), 99.8 (d, C-9), 99.7 (d, C-10), 99.6 (d, C-11), 99.5 (d, C-12), 99.4 (d, C-13), 99.3 (d, C-14), 99.2 (d, C-15), 99.1 (d, C-16), 99.0 (d, C-17), 98.9 (d, C-18), 98.8 (d, C-19), 98.7 (d, C-20), 98.6 (d, C-21), 98.5 (d, C-22), 98.4 (d, C-23), 98.3 (d, C-24), 98.2 (d, C-25), 98.1 (d, C-26), 98.0 (d, C-27), 97.9 (d, C-28), 97.8 (d, C-29), 97.7 (d, C-30), 97.6 (d, C-31), 97.5 (d, C-32), 97.4 (d, C-33), 97.3 (d, C-34), 97.2 (d, C-35), 97.1 (d, C-36), 97.0 (d, C-37), 96.9 (d, C-38), 96.8 (d, C-39), 96.7 (d, C-40), 96.6 (d, C-41), 96.5 (d, C-42), 96.4 (d, C-43), 96.3 (d, C-44), 96.2 (d, C-45), 96.1 (d, C-46), 96.0 (d, C-47), 95.9 (d, C-48), 95.8 (d, C-49), 95.7 (d, C-50), 95.6 (d, C-51), 95.5 (d, C-52), 95.4 (d, C-53), 95.3 (d, C-54), 95.2 (d, C-55), 95.1 (d, C-56), 95.0 (d, C-57), 94.9 (d, C-58), 94.8 (d, C-59), 94.7 (d, C-60), 94.6 (d, C-61), 94.5 (d, C-62), 94.4 (d, C-63), 94.3 (d, C-64), 94.2 (d, C-65), 94.1 (d, C-66), 94.0 (d, C-67), 93.9 (d, C-68), 93.8 (d, C-69), 93.7 (d, C-70), 93.6 (d, C-71), 93.5 (d, C-72), 93.4 (d, C-73), 93.3 (d, C-74), 93.2 (d, C-75), 93.1 (d, C-76), 93.0 (d, C-77), 92.9 (d, C-78), 92.8 (d, C-79), 92.7 (d, C-80), 92.6 (d, C-81), 92.5 (d, C-82), 92.4 (d, C-83), 92.3 (d, C-84), 92.2 (d, C-85), 92.1 (d, C-86), 92.0 (d, C-87), 91.9 (d, C-88), 91.8 (d, C-89), 91.7 (d, C-90), 91.6 (d, C-91), 91.5 (d, C-92), 91.4 (d, C-93), 91.3 (d, C-94), 91.2 (d, C-95), 91.1 (d, C-96), 91.0 (d, C-97), 90.9 (d, C-98), 90.8 (d, C-99), 90.7 (d, C-100), 90.6 (d, C-101), 90.5 (d, C-102), 90.4 (d, C-103), 90.3 (d, C-104), 90.2 (d, C-105), 90.1 (d, C-106), 90.0 (d, C-107), 89.9 (d, C-108), 89.8 (d, C-109), 89.7 (d, C-110), 89.6 (d, C-111), 89.5 (d, C-112), 89.4 (d, C-113), 89.3 (d, C-114), 89.2 (d, C-115), 89.1 (d, C-116), 89.0 (d, C-117), 88.9 (d, C-118), 88.8 (d, C-119), 88.7 (d, C-120), 88.6 (d, C-121), 88.5 (d, C-122), 88.4 (d, C-123), 88.3 (d, C-124), 88.2 (d, C-125), 88.1 (d, C-126), 88.0 (d, C-127), 87.9 (d, C-128), 87.8 (d, C-129), 87.7 (d, C-130), 87.6 (d, C-131), 87.5 (d, C-132), 87.4 (d, C-133), 87.3 (d, C-134), 87.2 (d, C-135), 87.1 (d, C-136), 87.0 (d, C-137), 86.9 (d, C-138), 86.8 (d, C-139), 86.7 (d, C-140), 86.6 (d, C-141), 86.5 (d, C-142), 86.4 (d, C-143), 86.3 (d, C-144), 86.2 (d, C-145), 86.1 (d, C-146), 86.0 (d, C-147), 85.9 (d, C-148), 85.8 (d, C-149), 85.7 (d, C-150), 85.6 (d, C-151), 85.5 (d, C-152), 85.4 (d, C-153), 85.3 (d, C-154), 85.2 (d, C-155), 85.1 (d, C-156), 85.0 (d, C-157), 84.9 (d, C-158), 84.8 (d, C-159), 84.7 (d, C-160), 84.6 (d, C-161), 84.5 (d, C-162), 84.4 (d, C-163), 84.3 (d, C-164), 84.2 (d, C-165), 84.1 (d, C-166), 84.0 (d, C-167), 83.9 (d, C-168), 83.8 (d, C-169), 83.7 (d, C-170), 83.6 (d, C-171), 83.5 (d, C-172), 83.4 (d, C-173), 83.3 (d, C-174), 83.2 (d, C-175), 83.1 (d, C-176), 83.0 (d, C-177), 82.9 (d, C-178), 82.8 (d, C-179), 82.7 (d, C-180), 82.6 (d, C-181), 82.5 (d, C-182), 82.4 (d, C-183), 82.3 (d, C-184), 82.2 (d, C-185), 82.1 (d, C-186), 82.0 (d, C-187), 81.9 (d, C-188), 81.8 (d, C-189), 81.7 (d, C-190), 81.6 (d, C-191), 81.5 (d, C-192), 81.4 (d, C-193), 81.3 (d, C-194), 81.2 (d, C-195), 81.1 (d, C-196), 81.0 (d, C-197), 80.9 (d, C-198), 80.8 (d, C-199), 80.7 (d, C-200), 80.6 (d, C-201), 80.5 (d, C-202), 80.4 (d, C-203), 80.3 (d, C-204), 80.2 (d, C-205), 80.1 (d, C-206), 80.0 (d, C-207), 79.9 (d, C-208), 79.8 (d, C-209), 79.7 (d, C-210), 79.6 (d, C-211), 79.5 (d, C-212), 79.4 (d, C-213), 79.3 (d, C-214), 79.2 (d, C-215), 79.1 (d, C-216), 79.0 (d, C-217), 78.9 (d, C-218), 78.8 (d, C-219), 78.7 (d, C-220), 78.6 (d, C-221), 78.5 (d, C-222), 78.4 (d, C-223), 78.3 (d, C-224), 78.2 (d, C-225), 78.1 (d, C-226), 78.0 (d, C-227), 77.9 (d, C-228), 77.8 (d, C-229), 77.7 (d, C-230), 77.6 (d, C-231), 77.5 (d, C-232), 77.4 (d, C-233), 77.3 (d, C-234), 77.2 (d, C-235), 77.1 (d, C-236), 77.0 (d, C-237), 76.9 (d, C-238), 76.8 (d, C-239), 76.7 (d, C-240), 76.6 (d, C-241), 76.5 (d, C-242), 76.4 (d, C-243), 76.3 (d, C-244), 76.2 (d, C-245), 76.1 (d, C-246), 76.0 (d, C-247), 75.9 (d, C-248), 75.8 (d, C-249), 75.7 (d, C-250), 75.6 (d, C-251), 75.5 (d, C-252), 75.4 (d, C-253), 75.3 (d, C-254), 75.2 (d, C-255), 75.1 (d, C-256), 75.0 (d, C-257), 74.9 (d, C-258), 74.8 (d, C-259), 74.7 (d, C-260), 74.6 (d, C-261), 74.5 (d, C-262), 74.4 (d, C-263), 74.3 (d, C-264), 74.2 (d, C-265), 74.1 (d, C-266), 74.0 (d, C-267), 73.9 (d, C-268), 73.8 (d, C-269), 73.7 (d, C-270), 73.6 (d, C-271), 73.5 (d, C-272), 73.4 (d, C-273), 73.3 (d, C-274), 73.2 (d, C-275), 73.1 (d, C-276), 73.0 (d, C-277), 72.9 (d, C-278), 72.8 (d, C-279), 72.7 (d, C-280), 72.6 (d, C-281), 72.5 (d, C-282), 72.4 (d, C-283), 72.3 (d, C-284), 72.2 (d, C-285), 72.1 (d, C-286), 72.0 (d, C-287), 71.9 (d, C-288), 71.8 (d, C-289), 71.7 (d, C-290), 71.6 (d, C-291), 71.5 (d, C-292), 71.4 (d, C-293), 71.3 (d, C-294), 71.2 (d, C-295), 71.1 (d, C-296), 71.0 (d, C-297), 70.

OY 714 CTGCGCAACATGCTGTAACCCCTGCTCTCTACAAAATAATCAAAAATTAATGCGGGCANGCT 773
 |||||
 DB 275 CTTGGCCACATGCTGTAACCCCTGCTCTCTACTAAACACAAAATTAATGCGAGGGGGCT 216
 |||||
 OY 774 GATCGGCGCTGATCCAGCTCTGCGGAGGCTGAGGAGGAGGAATCGCTTGAACCG 833
 |||||
 DB 215 GGTGGGCGCTGATCCAGCTCTGCGGAGGCTGAGGAGGAGGAATCGCTTGAACCG 156
 |||||
 OY 834 GAGAGCGAGGCTGATCCAGCTCTGCGGAGGCTGAGGAGGAGGAATCGCTTGAACCG 893
 |||||
 DB 155 GAGAGCGAGGCTGATCCAGCTCTGCGGAGGCTGAGGAGGAGGAATCGCTTGAACCG 96
 |||||
 OY 894 GCGAATTCGCTCTAAATAATTAATAA 928
 |||||
 DB 95 CCGAACTCGCTCTCAAAAAAAGAAAAA 61
 |||||

RESULT 9
 AL708218 690 bp mRNA linear EST 22-MAR-2002
 LOCUS DKFZp686M2350_r1 686 (synonym: hlc3) Homo sapiens cDNA clone
 DEFINITION DKFZp686M2350_5', mRNA sequence.
 AL708218
 ACCESSION AL708218.1 GI:19691573
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 690)
 AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Well, B. and Wiemann, S.
 TITLE EST (Ottenwaelder, B., Obermaier, B., Mewes, H.W., Well, B. and Wiemann, S.)
 JOURNAL Unpublished
 COMMENT Contact: Ottenwaelder B
 MPRS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by MediGenomix (Martinsried/Germany) within the cDNA
 sequencing consortium of the German Genome Project. No sl sequence
 available.
 This clone (DKFZp686M2350) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubneweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 FEATURES
 source
 1..690
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp686M2350"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="686 (synonym: hlc3)"
 /note="Vector: pTRIPLEX2; Site_1: sfIIA; Site_2: sfIIB;
 cDNA-collection"
 BASE COUNT 162 a 185 c 225 g 118 t
 ORIGIN
 Query Match 23.2%; Score 248.6; DB 9; Length 690;
 Best Local Similarity 86.2%; Pred. No. 8.8e+02;
 Matches 275; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 OY 605 TAGGCAATTTAAAGCATCAGGCGGGCGCATGCTCAGCCCTGTATCCACACTT 664
 |||||
 DB 28 TAGAGCCCTTGAACGAGCATGCGCGGGCGGTGCTACGCCCTGTATCGACACTT 87
 |||||
 OY 665 TAGAGCGCGAGCGCGCGCATCAGCTGAGTGGGAGTTGACACAGCCGCGGCAACA 724
 |||||
 DB 88 GGGAGGCGCGAGCGCGCATCAGCTGAGTGGGAGTTGACACAGCCGCGGCAACA 147
 |||||

OY 725 TGTGTAACCTGCTCTACAAAATAATCAAAAATTAATGCGGGCATGCTGCTCT 784
 |||||
 DB 148 CGGCGAACTCGCTCTACTAAATAATCAAAAATTAATGCGAGTGGCTGCTGCT 207
 |||||
 OY 765 GTGATCCAGCTACCTCGGAGGCTGAGGAGGAGGAATCGCTTGAACCGGGGAGG 844
 |||||
 DB 208 GTATCCAGCTACCTCGGAGGCTGAGGAGGAGGAATCGCTTGAACCGGGGAGGAGG 267
 |||||
 OY 845 TTGCAAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 904
 |||||
 DB 268 TTGCAAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 327
 |||||
 OY 905 TCTAAAAAATAATTAATAA 923
 |||||
 DB 328 TCTCAAGAAAGAAAGAAAAA 346
 |||||

RESULT 10
 BM460657 1136 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT_6421229 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532003
 DEFINITION 5', mRNA sequence.
 BM460657
 ACCESSION BM460657.1 GI:18509697
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1136)
 AUTHORS NIH-MGC http://mhc.mci.nih.gov/
 TITLE NIH-MGC
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM1214 row: j column: 04
 High quality sequence stop: 534.
 Location/Qualifiers
 1..1136
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5532003"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."
 BASE COUNT 253 a 399 c 241 g 243 t
 ORIGIN
 Query Match 23.2%; Score 248.6; DB 12; Length 1136;
 Best Local Similarity 86.2%; Pred. No. 6e+02;
 Matches 275; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 OY 605 TAGGCAATTTAAAGCATCAGGCGGGCGCATGCTCAGCCCTGTATCCACACTT 664
 |||||
 DB 134 TAGAGCCCTTGAACGAGCATGCGCGGGCGGTGCTACGCCCTGTATCGACACTT 193
 |||||
 OY 665 TAGAGCGCGAGCGCGCGCATCAGCTGAGTGGGAGTTGACACAGCCGCGGCAACA 724
 |||||
 DB 194 GGGAGGCGCGAGCGCGCATCAGCTGAGTGGGAGTTGACACAGCCGCGGCAACA 253
 |||||
 OY 725 TGTGTAACCTGCTCTACAAAATAATCAAAAATTAATGCGGGCATGCTGCTCT 784
 |||||


```

? CURRENT APPLICATION NUMBER: US/09/641,638
? CURRENT FILING DATE: 2000-08-16
? PRIOR APPLICATION NUMBER: US 09/502,330
? PRIOR FILING DATE: 2000-02-11
? PRIOR APPLICATION NUMBER: US 60/133,200
? PRIOR FILING DATE: 1999-05-07
? PRIOR APPLICATION NUMBER: US 09/275,267
? PRIOR FILING DATE: 1999-03-23
? PRIOR APPLICATION NUMBER: US 60/119,917
? PRIOR FILING DATE: 1999-02-12
? NUMBER OF SEQ ID NOS: 1304
? SOFTWARE: Patent.pm
? SEQ ID NO 316
? LENGTH: 1001
? TYPE: DNA
? ORGANISM: Homo Sapiens
? FEATURE:
? NAME/KEY: allele
? LOCATION: 501
? OTHER INFORMATION: 12-361-320 : polymorphic base G or T
? NAME/KEY: misc_binding
? LOCATION: 481..500
? OTHER INFORMATION: 12-361-320.misl, potential
? NAME/KEY: misc_binding
? LOCATION: 502..520
? OTHER INFORMATION: 12-361-320.mis2, complement
? NAME/KEY: primer_bind
? LOCATION: 182..200
? OTHER INFORMATION: upstream amplification primer
? NAME/KEY: primer_bind
? LOCATION: 677..696
? OTHER INFORMATION: downstream amplification primer, complement
? NAME/KEY: misc_binding
? LOCATION: 489..513
? OTHER INFORMATION: 12-361-320 potential probe
US-09-641-638-316

Query Match      44.7%; Score 22.8; DB 4; Length 1001;
Best Local Similarity 71.4%; Pred. No. 6.3;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy      5 CCTTTAAGGCTTGAATGTCGCAACTGTGATGTACT 46
Db      741 CCTTTTCGGCTTCATTTCCCATCTGCAATGTGACATT 782

RESULT 3
US-09-641-638-317
? Sequence 317, Application US/09641638
? Patent No. 6432648
? GENERAL INFORMATION:
? APPLICANT: Blumenfeld, Marta
? APPLICANT: Bougueleret, Lydie
? APPLICANT: Chumakov, Ilya
? APPLICANT: Cohen, Annick
? TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
? FILE REFERENCE: GENSET.05ICP1
? CURRENT APPLICATION NUMBER: US/09/641,638
? CURRENT FILING DATE: 2000-08-16
? PRIOR APPLICATION NUMBER: US 09/502,330
? PRIOR FILING DATE: 2000-02-11
? PRIOR APPLICATION NUMBER: US 60/133,200
? PRIOR FILING DATE: 1999-05-07
? PRIOR APPLICATION NUMBER: US 09/275,267
? PRIOR FILING DATE: 1999-03-23
? PRIOR APPLICATION NUMBER: US 60/119,917
? PRIOR FILING DATE: 1999-02-12
? NUMBER OF SEQ ID NOS: 1304
? SOFTWARE: Patent.pm
? SEQ ID NO 317
? LENGTH: 1001
? TYPE: DNA
```

```

? ORGANISM: Homo Sapiens
? FEATURE:
? NAME/KEY: allele
? LOCATION: 501
? OTHER INFORMATION: 12-361-388 : polymorphic base A or G
? NAME/KEY: misc_binding
? LOCATION: 482..500
? OTHER INFORMATION: 12-361-388.misl
? NAME/KEY: misc_binding
? LOCATION: 502..521
? OTHER INFORMATION: 12-361-388.mis2, potential complement
? NAME/KEY: primer_bind
? LOCATION: 114..132
? OTHER INFORMATION: upstream amplification primer
? NAME/KEY: primer_bind
? LOCATION: 609..628
? OTHER INFORMATION: downstream amplification primer, complement
? NAME/KEY: misc_binding
? LOCATION: 489..513
? OTHER INFORMATION: 12-361-388 potential probe
US-09-641-638-317

Query Match      44.7%; Score 22.8; DB 4; Length 1001;
Best Local Similarity 71.4%; Pred. No. 6.3;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy      5 CCTTTAAGGCTTGAATGTCGCAACTGTGATGTACT 46
Db      673 CCTTTTCGGCTTCATTTCCCATCTGCAATGTGACATT 714

RESULT 4
US-08-701-233B-2/C
? Sequence 2, Application US/08701233B
? Patent No. 5861308
? GENERAL INFORMATION:
? APPLICANT: Pfriundschuh, Michael; Renner, Christoph
? TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS ASSOCIATED WITH T CELL
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESSES:
? ADDRESS: Pelfe & Lynch
? STREET: 805 Third Avenue
? CITY: New York City
? STATE: New York
? ZIP: 10022
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
? COMPUTER: IBM
? OPERATING SYSTEM: PC-DOS
? SOFTWARE: Wordperfect
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/701,233B
? FILING DATE: 21-August-1996
? CLASSIFICATION: 536
? ATTORNEY/AGENT INFORMATION:
? NAME: Hanson, No. 5861308man D.
? REGISTRATION NUMBER: 30,946
? REFERENCE/DOCKET NUMBER: LUD 5440
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 688-9200
? TELEFAX: (212) 838-3884
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 380 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
US-08-701-233B-2

Query Match      42.0%; Score 21.4; DB 2; Length 380;
Best Local Similarity 63.3%; Pred. No. 17;
Matches 31; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
```


TITLE OF INVENTION: Theeocf
FILE REFERENCE: 5914-074-999
CURRENT APPLICATION NUMBER: US/09/186,188B
CURRENT FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 08/842,445
PRIOR FILING DATE: 1997-04-24
PRIOR APPLICATION NUMBER: 08/638,617
PRIOR FILING DATE: 1996-04-26
NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 26
LENGTH: 1094
TYPE: DNA
ORGANISM: Plant
US-09-186-188B-26

Query Match 41.6%; Score 21.2; DB 4; Length 1094;
Best Local Similarity 69.0%; Pred. No. 27;
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 10 TTAAGGCTTGATGTCGCACTGTCATGTACCTTAAG 51
DB 927 TTACGGTTTCCTTCTTCTACATCATTTGTACCTTACAG 886

RESULT 10
US-08-248-839C-9
Sequence 9, Application US/0824839C
Patent No. 5843702
GENERAL INFORMATION:
APPLICANT: McConnell, David
APPLICANT: Devine, Kevin
APPLICANT: O'Kane, Charles
TITLE OF INVENTION: A Gene Expression System
NUMBER OF SEQUENCES: 185
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 5843702disk of NO. 5843702th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,839C
FILING DATE: 25-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3614,214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 180..518
US-08-248-839C-9

Query Match 41.6%; Score 21.2; DB 2; Length 1201;
Best Local Similarity 76.5%; Pred. No. 28;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TGCTCCTTTAAGGCTTGATGTCGCACTGT 34
DB 64 TTCCTTTTAAACATCATGTATGTTGAACACTG 97

RESULT 11
US-09-080-855-17/c
Sequence 17, Application US/09080855A
Patent No. 6083721
GENERAL INFORMATION:
APPLICANT: Saras, Jan
APPLICANT: Franz, Petra
APPLICANT: Aspenstrm, Pontus
APPLICANT: Hellman, ulf
APPLICANT: Genez, Leonel Jorge
APPLICANT: Heldin, Carl-Henrik
TITLE OF INVENTION: PARC, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL
FILE REFERENCE: L0461/7030
CURRENT APPLICATION NUMBER: US/09/080,855A
CURRENT FILING DATE: 1998-05-18
EARLIER APPLICATION NUMBER: 08/805,583
EARLIER FILING DATE: 1997-02-25
NUMBER OF SEQ ID NOS: 39
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 458
TYPE: DNA
ORGANISM: Mus musculus
US-09-080-855-17

Query Match 41.2%; Score 21; DB 3; Length 458;
Best Local Similarity 73.0%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 8 TTTAAGGCTTGATGTCGCACTGTCATGTGTACA 44
DB 414 TTATGTGCTTGAGGCTGCATTGCTGTGTAGA 378

RESULT 12
US-09-566-076-17/c
Sequence 17, Application US/09566076
Patent No. 6475775
GENERAL INFORMATION:
APPLICANT: Saras, Jan
APPLICANT: Franz, Petra
APPLICANT: Aspenstrm, Pontus
APPLICANT: Hellman, ulf
APPLICANT: Genez, Leonel Jorge
APPLICANT: Heldin, Carl-Henrik
TITLE OF INVENTION: PARC, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL
FILE REFERENCE: L0461/7030
CURRENT APPLICATION NUMBER: US/09/566,076
CURRENT FILING DATE:
EARLIER APPLICATION NUMBER: 09/080,855
EARLIER FILING DATE: 1998-05-18
NUMBER OF SEQ ID NOS: 39
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 458
TYPE: DNA
ORGANISM: Mus musculus
US-09-566-076-17

Query Match 41.2%; Score 21; DB 4; Length 458;
Best Local Similarity 73.0%; Pred. No. 26;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 8 TTTAAGGCTTGATGTCGCACTGTCATGTGTACA 44
DB 414 TTATGTGCTTGAGGCTGCATTGCTGTGTAGA 378

RESULT 13

US-09-085-761A-45/7

Sequence 45; Application US/09085761A

Patent No. 6335178

GENERAL INFORMATION:

APPLICANT: Welner, Joel H.

APPLICANT: Turner, Raymond J.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN

TITLE OF INVENTION: SECRETION

NUMBER OF SEQUENCES: 77

CORRESPONDENCE ADDRESS:

ADDRESSEE: Modlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/085,761A

FILING DATE: 28-MAY-1998

CLASSIFICATION: 405

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Peter G.

REGISTRATION NUMBER: 32,847

REFERENCE/DOCKET NUMBER: DAB-03355

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8438

INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:

LENGTH: 4120 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-085-761A-45

Query Match 41.2% Score 21; DB 4; Length 4120;

Best Local Similarity 56.7%; Prod. No. 4;

Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 5 GCCTTAAAGCTTGAAGCTGCAATGTCATGTCAGCTTAA 49

DB 1507 GGTGATAGAGCTTGAAGCTTGAAGCTGTCATGTCAGCTTAA 1463

RESULT 14

US-09-053-197A-3/5

Sequence 3; Application US/09053197A

Patent No. 6022952

GENERAL INFORMATION:

APPLICANT: Welner, Joel H.

APPLICANT: Turner, Raymond J.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN

TITLE OF INVENTION: SECRETION

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Modlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SUMMARY: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/053,197A

FILING DATE: 01-APR-1998

CLASSIFICATION: 445

ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin L.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: DAB-04243

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8419

TELEFAX: (415) 397-8438

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 22108 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

US-09-053-197A-3

Query Match 41.2% Score 21; DB 3; Length 22108;

Best Local Similarity 56.7%; Prod. No. 74;

Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 5 GCCTTAAAGCTTGAAGCTGCAATGTCATGTCAGCTTAA 49

DB 17559 GGTGATAGAGCTTGAAGCTTGAAGCTGTCATGTCAGCTTAA 17815

RESULT 15

US-09-085-761A-4/6

Sequence 3; Application US/09085761A

Patent No. 6335178

GENERAL INFORMATION:

APPLICANT: Welner, Joel H.

APPLICANT: Turner, Raymond J.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN

TITLE OF INVENTION: SECRETION

NUMBER OF SEQUENCES: 77

CORRESPONDENCE ADDRESS:

ADDRESSEE: Modlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/085,761A

FILING DATE: 28-MAY-1998

CLASSIFICATION: 445

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Peter G.

REGISTRATION NUMBER: 32,847

REFERENCE/DOCKET NUMBER: DAB-03355

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8438

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 22108 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

US-09-085-761A-3

Query Match 41.2% Score 21; DB 4; Length 22108;

Best Local Similarity 66.7%; Pred. No. 73;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 5 CCTTTAAGGGCTTGATGTCTGCACTGTGATGTGACTTAA 49
| | | | | | | | | | | | | | | | | | | | | |
Db 17559 CGTGATTAAGCGGTTGAGTATCTCTACAGACATGTTTACGGTTPA 17515

Search completed: August 20, 2003, 14:32:41
Job time : 8.26518 secs

GenCore version 5.1.6
Copyright (c) 1995 - 2003 CompuGen Ltd.

0M nucleotide sequence search using SW model

Run on: August 20, 2003, 09:44:27 : Search time 20.0283 Seconds
(without alignments)
6871.827 Million cell updates/sec

File: us-09-761-466-2

Perfect score: 1
Sequence: 1
Identity: 100%

Scoring table: uscore 100, Gapext 1.0

Searched: 2552756 s-rqs, 149719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 7

Maximum DB seq length: 200000000

Post-processing: Minimum Match 9%

Listing first 45 summaries

Database:

N_Rnaseq-1370003.*
1: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL/ncdata/a/geneseq/geneseq-emb1/NA2003.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length (B)	HL	Description
1	51	100%	51	22	AAS0959
2	51	100%	137	22	DNA encoding mcsx/
3	51	100%	745	22	DNA encoding mcsx/
4	43.6	46.3	445	21	Genomic DNA #1 enc
5	23.2	45.5	419	21	Human prostate spe
6	23	45.1	28	14	DNA encoding mcsx/
7	22.8	44.7	191	21	Human prostate DNA
8	22.8	44.7	191	21	Human prostate DNA

9	22.8	44.7	1194	18	AAS2433
10	22.8	44.7	1431	24	AAS6833
11	22.8	44.7	13206	24	AAS14494
12	22.8	44.7	14209	24	AAS19943
13	22.8	44.7	15500	22	AAS1594
14	22.8	44.7	15500	22	AAS6607
15	22.8	44.7	15500	25	AAS74784
16	22.8	44.7	15500	25	AAS16684
17	22.8	44.7	15500	25	AAS67374
18	22.8	44.7	155199	25	AAS3456
19	22.8	44.3	257	25	AAS4402
20	22.6	44.3	257	25	AAS77186
21	22.6	44.3	790	20	AAS6979
22	22.4	43.9	653	24	AAS8119
23	22.4	43.9	1437	24	AAS8119
24	22.4	43.9	1620	22	AAS38421
25	22.4	43.9	2641	24	AAS24252
26	22.2	43.5	142	24	AAS74597
27	22.2	43.5	752	22	AAS4402
28	22.2	43.5	1265	20	AAS4591
29	22.2	43.5	2583	24	AAS2245
30	22.2	43.5	3634	24	AAS3587
31	22.2	43.5	3634	24	AAS5095
32	22.2	43.5	3635	24	AAS78105
33	22.2	43.5	3635	24	AAS94801
34	22.2	43.5	4635	25	AAS50143
35	22.2	43.5	5459	22	AAS2036
36	22.2	43.5	8400	22	AAS1604
37	22.2	43.1	512	21	AAS4473
38	22.2	43.1	1567	21	AAS7943
39	22.2	43.1	2951	22	AAS1795
40	22.2	43.1	4111	22	AAS795
41	22.2	43.1	167343	24	AAS4405
42	22.2	43.1	167343	24	AAS7279
43	21.8	42.7	408	16	AAS1769
44	21.8	42.7	1943	22	AAS455
45	21.8	42.7	1943	22	AAS6923

ALIGNMENTS

RESULT 1	AAS0959	standard; DNA: 51 BP.
ID	AAS0959	standard; DNA: 51 BP.
XX	AAS0959	
AC	AAS0959	
XX	AAS0959	
XX	21-OCT-2001 (first entry)	
DE	DNA encoding mcsx/NKX2.5 homology domain A2.	
XX		
KW	mcsx/NKX2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;	
KW	Therapeutic; heart tissue; gene therapy; mouse; US.	
XX		
OS	Mus musculus.	
XX		
FN	W200151006-A2.	
XX		
XX	15-JUN-2001.	
XX		
XX	16 JAN 2001: 200106-1501511.	
XX		
XX	14 JAN 2000: 200005-0176419.	
XX		
XX	(METH-) BETH ISRAEL DEACONESS MEDICAL CENT.	
XX		
XX	Age 1W. 12mo S.	
XX		
XX	WPI: 2001-451809/48.	
XX		
XX	New cardiac specific cell enhancer elements, useful for specifically	
XX	expressing gene in cardiac cell, as cardiac marker of cardiomyocyte	

PT induction, e.g. for optimizing cardiomyocyte induction -
XX
PS Claim 1; Fig 5A; 66pp; English.
XX The sequence represents the coding sequence of cardiac enhancer
CC mcsx/Nkx2.5 homology domain A2. The nucleic acid is useful for
CC specifically expressing a gene in a cardiac cell, as an earlier marker
CC of cardiomyocyte induction, e.g. for optimizing cardiomyocyte induction.
CC Genes expressed in the cardiac cell-specific manner are useful for the
CC targeted expression of genes encoding therapeutic proteins for the
CC treatment of damaged heart tissue. Cardiac specific enhancer elements may
CC be used for gene therapy.
XX
SQ Sequence 51 BP; 12 A; 10 C; 11 G; 18 T; 0 other;
Query Match 100.0%; Score 51; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCTCCTTTTAAGGGCTTGATGTCGCACTGTCATGTGTACCTTAAG 51
DB 1 TGCTCCTTTTAAGGGCTTGATGTCGCACTGTCATGTGTACCTTAAG 51
RESULT 2
AAS09960 standard; DNA; 1072 BP.
XX
AC AAS09960:
XX
DT 24-OCT-2001 (first entry)
XX
DE DNA encoding mcsx/Nkx2.5 homology domains A1 + A2, and intervening DNA.
XX
KM mcsx/Nkx2.5 homology domain; cardiac enhancer; cardiac cell;
KW cardiomyocyte induction; therapeutic; heart tissue; gene therapy;
KW mouse; ds.
XX
OS Mus musculus.
XX
PN WO200151006-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01511.
XX
PR 14-JAN-2000; 2000US-0176419.
XX
(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lee IW, Izumo S;
XX
DR WPI; 2001-451809/48.
XX
PT New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
PS Claim 1; Fig 5B; 66pp; English.
XX
SQ The sequence represents the coding sequence of cardiac enhancer
CC mcsx/Nkx2.5 homology domains A1 + A2 and intervening sequence. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
SQ Sequence 1072 BP; 273 A; 290 C; 297 G; 212 T; 0 other;
Query Match 100.0%; Score 51; DB 22; Length 1072;
Best Local Similarity 100.0%; Pred. No. 6.6e-10;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCTCCTTTTAAGGGCTTGATGTCGCACTGTCATGTGTACCTTAAG 51
DB 1022 TGCTCCTTTTAAGGGCTTGATGTCGCACTGTCATGTGTACCTTAAG 1072
RESULT 3
AAS09961 standard; DNA; 7836 BP.
XX
AC AAS09961:
XX
DT 24-OCT-2001 (first entry)
XX
DE Genomic DNA #1 encoding human Csx/Nkx2.5.
XX
KW Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
KW therapeutic; heart tissue; gene therapy; human; ds.
XX
OS Homo sapiens.
XX
PN WO200151006-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01511.
XX
PR 14-JAN-2000; 2000US-0176419.
XX
(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lee IW, Izumo S;
XX
DR WPI; 2001-451809/48.
XX
PT New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
PS Disclosure; Fig 4A; 66pp; English.
XX
SQ The sequence represents the genomic sequence #1 of human Csx/Nkx2.5. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
SQ Sequence 7836 BP; 2164 A; 1938 C; 1907 G; 1825 T; 2 other;
Query Match 100.0%; Score 51; DB 22; Length 7836;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCTCCTTTTAAGGGCTTGATGTCGCACTGTCATGTGTACCTTAAG 51
DB 4609 TGCTCCTTTTAAGGGCTTGATGTCGCACTGTCATGTGTACCTTAAG 4659
RESULT 4
ABN87717 standard; cDNA; 865 BP.
XX
AC ABN87717:
XX
DT 08-AUG-2002 (first entry)
XX
DE Human prostate specific gene cDNA sequence SEQ ID NO:68.
XX
KW Human; prostate specific gene; prostate specific protein; PSG; PSP;
KW prostate cancer; gene; ss.

XX	Homo sapiens.
XX	W20025600 A2.
PX	19-MAY-2002.
XX	05-NOV-2001; 2001W-0847243
FX	04-NOV-2000; 2000US-2457401.
XX	(DIND-) DIADENUS INC.
XX	Sun Y, Renjona H, Chen S, Liu C;
XX	WP1: 2002-471506/50.
XX	New prostate-specific nucleic acids and polypeptides, useful for
PT	identifying, diagnosing, monitoring, staging, imaging, and treating
PT	prostate cancer and non-cancerous disease states in prostate tissue -
XX	Claim 1; Page 187; 254pp; English.
XX	ABN87650 to ABN87789 represent human prostate-specific nucleic acids (I),
CC	and ABN79192 to ABN79295 represent human prostate-specific proteins (II)
CC	from the present invention. (I) and (II) have cytostatic activity. (I)
CC	can be used in gene therapy. The prostate-specific nucleic acids,
CC	polypeptides and compositions from the present invention can be used for
CC	identifying, diagnosing, monitoring, staging, imaging, and treating
CC	prostate cancer and non-cancerous disease states in prostate tissue; for
CC	identifying prostate tissue, for monitoring, identifying and/or designing
CC	academics and anti-oncologists of the polypeptides; in gene therapy; in
CC	producing transgenic animals and cells; for producing engineered prostate
CC	tissue for treatment and research; and as elements in an array or
CC	computer program for pattern recognition of prostate disorders. The
CC	nucleic acids may be used as hybridisation probes to detect, characterise
CC	and quantify hybridising nucleic acids in, and isolate hybridising
CC	nucleic acids from, both genomic and transcript-derived nucleic acid
XX	samples.
XX	Sequence 865 BP; 130 A; 126 G; 130 Q; 169 I; 250 Other;
XX	Query Match 46.3%; Score 23.6; PB 24; Length 865;
XX	Best Local Similarity 76.5%; Pval No. 25;
XX	Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0
OY	6 CTTTAAAGGCGCTGAATGTGTGCACCTGTATSTGAC 43
Dh	
Dh	240 CTTTAAAGGCAATTCATTCTCAACTTAATATGTSAC 277
RESEUL 5	AASB8398/c
Id	AASB8398 standard; cMA: 3400 BP.
XX	AASB8398:
XX	13-FEB-2002 (first entry)
DE	DNA encoding novel human diagnostic protein #24202.
KM	Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM	food supplement; medical; imaging; diagnosis; genetic disorder; ss.
XX	Homo sapiens.
XX	W0200175067-A2.
XX	11-OCT-2001.
XX	30-MAR-2001; 2001WO-05066A1.
XX	31-MAR-2000; 2000GS-154217.

1R	23-A03-2000: 2000US-06494167.	XX
XX	(HSE-) HYSEQ INC.	XX
PA	lormanac RT, Liu C, Tang YF;	XX
PI	Wp1: 2001-639362/73.	XX
DR	P-PSDB: AB024211.	XX
PI	New isolated polynucleotide and encoded polypeptides, useful in	PI
PI	diagnostics, for genetic disorders or other traits and to assess	PI
PI	responsibility for genetic disorders or other traits and to assess	PI
PI	biocompatibility	PI
XX	Claim 1: SEQ ID No 24202: 103bp; English.	XX
PS	The invention relates to isolated polynucleotide (I) and	PS
XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,	XX
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome	CC
CC	gene mapping, and in recombinant production of (II). The	CC
CC	polynucleotides are also used in diagnostics as expressed sequence tags	CC
CC	for identifying expressed genes. (II) is useful in gene therapy techniques	CC
CC	to restore normal activity of (II) or to treat disease states involving	CC
CC	(II). (II) is useful for generating antibodies against it, detecting or	CC
CC	quantitating a polypeptide in tissue, as molecular weight markers and as	CC
CC	a food supplement. (II) and its binding partners are useful in medical	CC
CC	diagnosis of sites expressing (II). (I) and (II) are useful for treating	CC
CC	diseases involving aberrant protein expression or biological activity.	CC
CC	The polypeptide and polynucleotide sequences have applications in	CC
CC	diagnostics, for genetic disorders or other traits to assess biocompatibility	CC
CC	responsible for genetic disorders or other traits to assess biocompatibility	CC
CC	and to produce other types of data and products dependent on DNA and	CC
CC	amino acid sequences. AAS64197-AAS94564 represent novel human	CC
CC	diagnostic coding sequences of the invention.	CC
CC	Note: The sequence data for this patent did not appear in the printed	CC
CC	specification, but was obtained in electronic format directly from WIPO	CC
CC	at http://wipo.int/pub/published_pat_sequences .	CC
XX	Sequence 3480 bp; 850 A; 1003 C; 946 G; 640 T; 1 other:	XX
XX	Query Match 45.5%; Score 23.27; DB 23; Length 3480;	XX
XX	Best Local Similarity 70.5%; Fred. No. 50;	XX
XX	Matches 31; Conservative 0; Mismatches 13; Indels 9; Caps 0;	XX
XX	5 CCTTTAGGCGCTTGAATGCTGTAACATGATCATGTCATTA 48	XX
XX		XX
XX	1321 CCTTGAAGGCGCTGATGCTTGTGCGCTTTATGTCGACATTA 1778	XX
XX	Result 6	XX
XX	AAV26023 standard; DNA; 528 bp.	XX
XX	AAV26023:	XX
XX	28-AUG-1998 (first entry)	XX
XX	Bacteriophage DNA SEQ ID No: 45.	XX
XX	Bacteriophage; pathogen; Escherichia coli 0157; bacteriophage; food;	XX
XX	sterilisation; enteric haemorrhage; ds	XX
XX	Bacteriophage.	XX
XX	W09808944-A1.	XX
XX	05-MAR-1998.	XX
XX	25-AUG-1997; 97W0-JP02957.	XX
XX	19-APR-1997; 97JP-0135716.	XX
XX	26-AUG-1996; 96JP-0261132.	XX
XX	14-APR-1997; 97JP-0140246.	XX

```
XX (BIOV-) BIO VENTURE BANK CO LTD.
PA
XX Nishikori K, Takahashi S.
PI
XX WPI; 1998-230262/20.
DR
XX Bacteriophage with high specificity for particular pathogens such as
PT Escherichia coli 0157 - is incorporated in bactericides for food
PT sterilisation
XX
XX Claim 3; Page 45; 54pp; Japanese.
XX
CC The present sequence represents a DNA sequence from a novel
CC bacteriophage which has high specificity for pathogenic bacteria,
CC especially for strains of Escherichia coli causing enteric haemorrhage,
CC such as E. coli 0157. The bacteriophage can be incorporated into
CC bio-bactericidal compositions. These can contain more than one
CC bacteriophage strain in order to have a bactericidal effect against more
CC than one pathogen at the same time. The bio-bactericide may also contain
CC an amino-acid or other material to control the pH to 6.5-7.5 and ensure
CC stability of the bacteriophage, such as glycine, arginine or lysine.
CC The bio-bactericidal composition can be used for treating food (such as
CC meat or fish) to prevent bacterial contamination, e.g. in fresh food or
CC in the kitchens of restaurants, schools and other institutions. It may
CC be formulated as a spray. It can also be used to sterilise working
CC surfaces, aprons and other clothing. The bio-bactericides are safe to
CC humans but very potent against pathogenic bacteria.
SQ Sequence 528 BP; 171 A; 89 C; 129 G; 137 T; 2 other;

Query Match 45.1%; Score 23; DB 19; Length 528;
Best Local Similarity 72.5%; Pred. No. 37;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 11 AAGGCTTGAAATGCTGCAACTGTCATGTACACTTAA 50
    ||| ||||| ||||| || ||| ||| |||
DB 312 AAGAGCTTGAAATGCTGCAAGGTACATGTCGACCTCANA 351

RESULT 7
AAC57682
ID AAC57682 standard; DNA: 1001 BP.
XX
AC AAC57682;
XX
DT 25-JAN-2001 (first entry)
XX
DE Arachidonic acid metabolism related genomic biallelic marker #316.
XX
KW Human: biallelic marker; arachidonic acid metabolism; genotyping;
KW detection: hybridisation; phenotype; haplotype; SNP; polymorphic base;
KW single nucleotide polymorphism; hybridisation assay; sequencing assay;
KW specific amplification assay; identification; ERM; 12-LO-RBM;
KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
XX
XX Homo sapiens.
OS
XX WO200047771-A2.
PN
XX 17-AUG-2000.
PD
XX
PF 11-FEB-2000; 2000WO-IB00184.
PE
XX
PR 12-FEB-1999; 99US-0119917.
PR 23-MAR-1999; 99US-0275267.
PR 07-MAY-1999; 99US-0133200.
XX
XX (GEST ) GENSET.
PA
XX Blumenfeld M, Bougueleret L, Chumakov I;
PI
XX WPI; 2000-571881/53.
DR
```

```
XX Novel biallelic markers useful for detecting conditions and genotypes
PT associated with arachidonic acid metabolism -
PT
XX Claim 13; Page 501-502; 802pp; English.
PS
XX
XX The present invention describes polynucleotides including biallelic
CC markers derived from genes involved in arachidonic acid metabolism and
CC from genomic regions flanking those genes. Methods from the present
CC invention may be used to select individuals for clinical trials and
CC predict responses to treatment with drugs. The polynucleotides may be
CC used in hybridisation assays, sequencing assays and specific
CC amplification assays for identifying an eicosanoid-related biallelic
CC segment (ERBM) or 12-LO-related biallelic marker, and for amplifying a
CC segment of nucleotides containing an ERBM. The polynucleotides are
CC useful in diagnostic kits. The markers may be used to detect conditions
CC and genotypes associated with arachidonic acid metabolism. AAC5367 to
CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the
CC exemplification of the present invention.
CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as
CC SNPs) in the polynucleotide sequences from the present invention have
CC been given as their corresponding degenerate bases e.g. a polymorphic
CC base of C or T has been given as Y.
XX
SQ Sequence 1001 BP; 187 A; 264 C; 272 G; 277 T; 1 other;

Query Match 44.7%; Score 22.8; DB 21; Length 1001;
Best Local Similarity 71.4%; Pred. No. 52;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 5 CCTTTAAGGGCTTGAAATGCTGCAACTGTCATGTGACT 46
    ||||| ||||| ||| || ||| ||||| |||
DB 741 CCTTTTCTTGCTTCATTCATTCCTCCATCTGCAATGTGACACT 782

RESULT 8
AAC57683
ID AAC57683 standard; DNA: 1001 BP.
XX
AC AAC57683;
XX
DT 25-JAN-2001 (first entry)
XX
DE Arachidonic acid metabolism related genomic biallelic marker #317.
XX
KW Human: biallelic marker; arachidonic acid metabolism; genotyping;
KW detection: hybridisation; phenotype; haplotype; SNP; polymorphic base;
KW single nucleotide polymorphism; hybridisation assay; sequencing assay;
KW specific amplification assay; identification; ERM; 12-LO-RBM;
KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
XX
XX Homo sapiens.
OS
XX WO200047771-A2.
PN
XX 17-AUG-2000.
PD
XX
PF 11-FEB-2000; 2000WO-IB00184.
PE
XX
PR 12-FEB-1999; 99US-0119917.
PR 23-MAR-1999; 99US-0275267.
PR 07-MAY-1999; 99US-0133200.
XX
XX (GEST ) GENSET.
PA
XX Blumenfeld M, Bougueleret L, Chumakov I;
PI
XX WPI; 2000-571881/53.
DR
XX Novel biallelic markers useful for detecting conditions and genotypes
PT associated with arachidonic acid metabolism -
PT
XX Claim 13; Page 502; 802pp; English.
PS
```


XX The present invention describes polynucleotides including allelic
 CC markers derived from genes involved in arachidonic acid metabolism and
 CC from genomic regions flanking those genes. Methods from the present
 CC invention may be used to select individuals for clinical trials and
 CC predict responses to treatment with drugs. The polynucleotides may be
 CC used in hybridisation assays, sequencing assays and specific
 CC amplification assays for identifying an eicosanoid-related allelic
 CC marker (ERBM) or 12-O-related allelic marker, and for amplifying a
 CC segment of nucleotides containing an ERBM. The polynucleotides are
 CC useful in diagnostic kits. The markers may be used to detect conditions
 CC and genotypes associated with arachidonic acid metabolism. AAC57367 to
 CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the
 CC exemplification of the present invention.
 CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as
 CC SNPs) in the polynucleotide sequences from the present invention have
 CC been given as their corresponding degenerate bases e.g. a polymorphic
 CC base of C or T has been given as Y.

Sequence 1001 BP: 181 A: 269 G: 291 T: 1 other;

Query Match 44.7%; Score 22.8; DB 21; Length 1001;

Best Local Similarity 71.4%; Pred. No. 52;

Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

5 GCCTTAACGACCTTCAATGCTGCACTGCACTGACACT 46

673 GCTTTTCTGCTTCAATTTGGCATTTGCAATGTCACAT 714

RESULT 9

AAV74433/C

ID AAV74433 standard; DNA: 1194 BP.

AAV74433;

16-MAR-1999 (first entry)

Staphylococcus aureus could SEQ ID #122.

Computer readable medium; vaccine; S. aureus infection; immunodetection;
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 skin infection; surgical wound infection; scalded skin syndrome;
 toxic shock syndrome; ds.

Staphylococcus aureus.

EP766519-A2.

40-JUL-1997.

07-JAN-1997; 97EP-0100117.

05-JAN-1996; 96US-0009661.

(HUMAN-) HUMAN GENOME SCI. INC.

Barash SC, Choi CH, Dillon PJ, Fannon MR, Kusch CA;

Rosen CA;

WPI: 1997-37922/15.

Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 stored on computer readable medium and used in the production of
 anti-S. aureus vaccines

Claim 1: Page 684: 327pp; Encl151.

This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 of the invention. The DNA sequences are reported on a computer readable
 medium, preferably selected from a floppy or hard disk, random access
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 the *S. aureus* DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against *S. aureus* infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the *S. aureus* DNA sequences contained on the
 CC computer readable medium.

Sequence 1194 BP: 429 A: 158 C: 221 G: 383 T: 3 other;

Query Match 44.7%; Score 22.8; DB 16; Length 1194;

Best Local Similarity 66.0%; Pred. No. 54;

Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

2 GCTCTTTTAAAGGTTTCAATGCTGCACTGCACTGACACT 51

661 GCTACTGCAATGCTGCACTGCACTGCACTGCACTGCACT 812

RESULT 10

ABN68333/C

ID ABN68333 standard; DNA: 1431 BP.

ABN68333;

01-JUL-2002 (first entry)

Streptococcus polynucleotide SEQ ID NO 4579.

Streptococcus GAS; GAS; group B *Streptococcus*; *Streptococcus* *ausarum*;
 group A *Streptococcus*; *Streptococcus* pyogenes; *actinobacillus* *group*;
 antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

Streptococcus pyogenes.

WO2002/4771-A2.

02-MAY-2002.

29-06T-2001; 2001WO-0804789.

27-06T-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0026727.

07-MAR-2001; 2001GB-0009540.

(CHIR-) CHIRON SPA

(GENO-) INST GENOMIC RES.

Telford J, Mastignani V, Margarit Kos Y, Grandi G, Fraser G;

Tettein H;

New *Streptococcus* protein for the treatment or prevention of infection
 or disease caused by *Streptococcus* bacteria, such as meningitis, and
 for detecting a compound that binds to the protein -

Claim 7: Page 3620: 4525pp; Encl151.

The invention relates to a protein (ABP25413-ABP30895) from group B
Streptococcus (GAS) (*Streptococcus* *ausarum*) or group A *Streptococcus* (*S.*
Streptococcus *pyogenes*), comprising one of 5483 sequences (S1), given in
 the specification. The proteins have antibacterial and antiinflammatory
 activity. (1), nucleic acids encoding (1), ABP25413 ABP30895 and
 antibodies that bind (1) are used in the manufacture of reagents for
 the treatment or prevention of infection or disease caused by

[illegible]

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
DR
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX
PS Disclosure; SEQ ID NO 8171; 1701bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful.
CC In the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIFO at ftp.wifo.int/pub/published_pct_sequences.
XX
SQ Sequence 15500 BP; 3660 A; 4223 C; 4207 G; 3410 T; 0 other;
Query Match 44.7%; Score 22.8; DB 22; Length 15500;
Best Local Similarity 66.0%; Pred. No. 1e+02;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
OY 1 TGCTCCTTTTAAGCGCTTGAATGTCGCAACTGTCATGTGTACATTAA 50
DB 9272 TACTACTGTGACGATATTTTAAAGATTTCCTACTGTCATATATTCATTAA 9321
RESULT 14
AAS36609/c
ID AAS36609 standard; DNA: 15500 BP.
XX
AC AAS36609;
XX
DE Human cardiovascular system antigen genomic DNA SEQ ID No 2109.
XX
XX 17-DEC-2001 (first entry)
XX
XX Human cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotopic; dog;
KW antileukemic; antiproliferative; cytostatic; cardiact; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; vincinoid; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular infection; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;

KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX
XX
OS Homo sapiens.
PN WO200155321-A2.
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01340.
PE
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0198874.
PR 17-MAR-2000; 2000US-0198076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 18-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227099.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228927.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.

[illegible]

REFERENCE 2 (bases 1 to 81512)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 81512)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 81512)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 5 (bases 1 to 81512)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Jan 28, 2003 this sequence version replaced gi:20376952.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.3% of Sequence;
Estimated Total Number of Errors is 0.4.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 139,5b). It is clipped at the overlaps with AC008378 and AC106731. The number of bases overlapped with AC008378 is 4442 bps and with AC106731 is 15413 bps.

FEATURES
source
1..81512
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-281H14"
Location/Qualifiers
BASE COUNT 20011 a 19534 c 19980 g 21987 t
ORIGIN

Query Match 100.0%; Score 51; DB 9; Length 81512;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCTCCTTTAAGGGCTGAATGCTGCACTGTCATGTGACCTTAAG 51
|||||
Db 53141 TGCTCCTTTAAGGGCTGAATGCTGCACTGTCATGTGACCTTAAG 53091
|||||

RESULT 2
AL669924/c 137345 bp DNA linear PRI 24-APR-2002
LOCUS Human DNA sequence from clone Xkbac-2L19 on chromosome 6, complete sequence.
ACCESSION AL669924
VERSION AL669924.9 GI:20068712
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1
AUTHORS Peck, A.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19572797.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from a CHORI-501 human bac - PGF cell line library VECTOR: pTARBAC2.1
This sequence was generated from part of bacterial clone contigs constructed by the MHC Haplotype Consortium and collaborators. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6/MHC>.

FEATURES
source
1..137345
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Xkbac-2L19"
/clone_lib="CHORI-501"
Location/Qualifiers
BASE COUNT 33377 a 33207 c 33325 g 37436 t
ORIGIN

Query Match 100.0%; Score 51; DB 9; Length 137345;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCTCCTTTAAGGGCTGAATGCTGCACTGTCATGTGACCTTAAG 51
|||||
Db 53367 TGCTCCTTTAAGGGCTGAATGCTGCACTGTCATGTGACCTTAAG 53317
|||||

RESULT 3
AC144621 158806 bp DNA linear HTG 06-MAY-2003
LOCUS Mus musculus chromosome UNK clone RP24-335114, WORKING DRAFT
DEFINITION SEQUENCE, 69 unordered pieces.
ACCESSION AC144621
VERSION AC144621.1 GI:30387755
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 158806)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone unpublished
JOURNAL 2 (bases 1 to 158806)
REFERENCE McPherson, J.D. and Waterston, R.H.
AUTHORS Direct Submission
TITLE Submitted (06-MAY-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
JOURNAL

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu
Project Information
Center project name: M.BB0335114
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%


```
Chemistry: Dye-Primer 51: 0% of reads
Chemistry: Dye-Terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 14344 bases at least Q40
Consensus quality: 14344 bases at least Q30
Consensus quality: 14344 bases at least Q20
Insert size: 217000: aaurose-1p
Insert size: 153245: sum-of-contigs
Quality coverage: 1.93 in Q20 bases: sum-of-contigs
Quality coverage: 2.95 in Q20 bases: sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N. But the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1262: contig of 1262 bp in length
* 1263: gap of unknown length
* 1363 2803: contig of 1447 bp in length
* 2810 2909: gap of unknown length
* 2910 4113: contig of 1204 bp in length
* 4114 4213: gap of unknown length
* 4214 5492: contig of 1279 bp in length
* 5493 5592: gap of unknown length
* 5593 5902: contig of 1310 bp in length
* 5903 7052: gap of unknown length
* 7053 8255: contig of 1253 bp in length
* 8256 8356 9423: gap of unknown length
* 9424 9523: gap of unknown length
* 9524 10777: contig of 1254 bp in length
* 10778 10877: gap of unknown length
* 10879 12123: contig of 1146 bp in length
* 12024 12124 13409: gap of 1286 bp in length
* 13410 13509: gap of unknown length
* 13510 15253: contig of 1744 bp in length
* 15254 15353: gap of unknown length
* 15354 16499: contig of 1296 bp in length
* 16500 16750 18223: contig of 1480 bp in length
* 18224 18323: gap of unknown length
* 18324 19662: contig of 1337 bp in length
* 19663 19762: gap of unknown length
* 19763 21124: contig of 1262 bp in length
* 21125 22467: contig of 1343 bp in length
* 22468 22567: gap of unknown length
* 22568 23615: contig of 1048 bp in length
* 23616 23715: gap of unknown length
* 23716 25451: contig of 1936 bp in length
* 25452 25751: gap of unknown length
* 25752 27281: contig of 1530 bp in length
* 27282 27381: gap of unknown length
* 27382 29420: contig of 2039 bp in length
* 29421 29520: gap of unknown length
* 29521 40301: contig of 1381 bp in length
* 40302 41301: gap of unknown length
* 41302 42159: contig of 1159 bp in length
* 42160 42459: gap of unknown length
* 42460 43786: contig of 1529 bp in length
* 43787 43886: gap of unknown length
* 43887 45134: contig of 1247 bp in length
* 45135 45234: gap of unknown length
* 45235 46486: contig of 1254 bp in length
* 46487 46586: gap of unknown length
* 46587 48485: contig of 1799 bp in length
* 48486 48585: gap of unknown length
* 48586 49714: contig of 1229 bp in length
* 49715 49814: gap of unknown length
*
* 49815 41078: contig of 1264 bp in length
* 41079 41178: gap of unknown length
* 41179 42273: contig of 1095 bp in length
* 42274 42373: gap of unknown length
* 42374 44223: contig of 1750 bp in length
* 44224 44323: gap of unknown length
* 44324 45908: contig of 1645 bp in length
* 45909 46008: gap of unknown length
* 46009 47545: contig of 1547 bp in length
* 47546 47645: gap of unknown length
* 47646 50012: contig of 2367 bp in length
* 50013 50112: gap of unknown length
* 50113 51978: contig of 1866 bp in length
* 51979 52078: gap of unknown length
* 52079 54113: contig of 2935 bp in length
* 54114 54213: gap of unknown length
* 54214 55585: contig of 1374 bp in length
* 55587 55687 57405: contig of 1719 bp in length
* 57406 57505: gap of unknown length
* 57506 58790: contig of 1281 bp in length
* 58791 58890: gap of unknown length
* 58891 60453: contig of 1564 bp in length
* 60454 62139: contig of 1596 bp in length
* 62150 62439: gap of unknown length
* 62440 63911: contig of 1661 bp in length
* 63912 64010: gap of unknown length
* 64011 65910: contig of 1900 bp in length
* 65911 66010: gap of unknown length
* 66011 67754: contig of 1744 bp in length
* 67755 67854: gap of unknown length
* 67855 69859: contig of 2005 bp in length
* 69860 71962: contig of 2003 bp in length
* 71963 72062: gap of unknown length
* 72063 74371: contig of 2409 bp in length
* 74372 74471: gap of unknown length
* 74472 76499: contig of 2028 bp in length
* 76500 78329: contig of 1730 bp in length
* 78330 78429: gap of unknown length
* 78430 81033: contig of 2604 bp in length
* 81034 81133: gap of unknown length
* 81134 82450: contig of 1317 bp in length
* 82451 82550: gap of unknown length
* 82551 85880: contig of 3330 bp in length
* 85881 85980: gap of unknown length
* 85981 89959: contig of 3979 bp in length
* 89960 90059: gap of unknown length
* 90060 92499: contig of 2440 bp in length
* 92500 92599: gap of unknown length
* 92600 95472: contig of 2774 bp in length
* 95473 98796: contig of 3324 bp in length
* 98797 98896: gap of unknown length
* 98897 101619: contig of 2724 bp in length
* 101620 104543: contig of 3744 bp in length
* 104544 105543: gap of unknown length
* 105544 108620: contig of 2957 bp in length
* 108621 108620: gap of unknown length
* 108621 111713: contig of 3094 bp in length
* 111714 111814: gap of unknown length
* 111815 115701: contig of 3887 bp in length
* 115702 115801: gap of unknown length
* 115802 117201: contig of 1919 bp in length
* 117202 117820: gap of unknown length
* 117821 121279: contig of 4459 bp in length
* 121280 121479: gap of unknown length
* 121480 125489: contig of 4110 bp in length
* 125490 125589: gap of unknown length
* 125590 128897: contig of 3308 bp in length
```

```

* 128898 12897: gap of unknown length
* 12898 13134: contig of 4137 bp in length
* 13135 13324: gap of unknown length
* 13325 13740: contig of 4168 bp in length
* 13743 13750: gap of unknown length
* 13753 14264: contig of 5140 bp in length
* 14263 14274: gap of unknown length
* 14273 14852: contig of 5810 bp in length
* 14853 14853: gap of unknown length
* 14853 15806: contig of 10154 bp in length.
* 15806: location/Qualifiers
  1. 15806
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="X"
    /clone="RP24-335114"
  1. 1262
    /note="assembly_name:Contig10"
  1363. 2809
    /note="assembly_name:Contig22"
  2910. 4113
    /note="assembly_name:Contig34"

misc_feature
  /note="assembly_name:Contig10"
  1363. 2809
  /note="assembly_name:Contig22"
  2910. 4113
  /note="assembly_name:Contig34"

misc_feature
  /note="assembly_name:Contig34"

Query Match 78.0%; Score 39.8; DB 2; Length 158806;
Best Local Similarity 86.3%; Pred. No. 5.7e-05;
Matches 44; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 TGCTCTTTTAAAGGCTGATGCTGCAACGTGTCATGTCACCTTAAG 51
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139093 TGCTCTTTTAAAGGCTGATGCTGCAACGTGTCATGTCACCTTAAG 139143

RESULT 4
LOCUS AC119699/c 171110 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-466D21, WORKING DRAFT SEQUENCE, 5
unorderd pieces.
ACCESSION AC119699.8 GI:24941740
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 171110)
Muzny,D.,Marle,,Metzker,M.,Lee,,Abramson,S.,Adams,C.,Alder,J.,
Allen,C.,Allen,H.,Alsbrooks,S.,Amin,A.,Angulano,D.,
Anyadebech,V.,Ayagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,
Baldwin,D.,Bandaraike,D.,Barber,M.,Barnstead,M.,Benahmed,F.,
Biswal,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,
Bryant,N.,Buhay,C.,Burch,P.,Burrell,K.,Calderon,E.,
Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,
Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,Diederich,D.,
Davila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Diya,K.,
Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Duval,B.,Eaves,K.,
Draper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,
Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,
Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P.,
Fraser,C.M.,Gabisi,A.,Ganta,R.,Grady,M.,Guerra,W.,Guevara,W.,
Gebregorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,
Gumarate,P.,Haaland,W.,Hamill,C.,Hamilton,C.,Hamilton,K.,
Harvey,Y.,Havlak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,
Hollins,B.,Howells,S.,Huiy,S.,Hume,J.,Idlebird,D.,Jaisvel,A.,
Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Johnson,A.,
Karpachy,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Kovar,C.,
Kowis,C.,Kraft,C.L.,Lebow,H.,Levan,D.,Lewis,L.,Li,Z.,Liu,J.,
Liu,J.,Liu,W.,Liu,Y.,London,P.,Longacre,S.,Lopez,J.,
Lorensbawa,L.,Loulsegad,H.,Lozado,R.J.,Lu,X.,Ma,J.,
Maheshwari,M.,Mahindartne,M.,Mahmoud,M.,Malloy,K.,Mangum,A.,

```

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Morgan,M., Morris,K., Morris,S., Munday,M., Murphy,M., Nait,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankwelu,O., Okwunonu,G., Olarnpungoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Polindexter,A., Popovic,D., Prims,E., Pu,L.,
Puzo,M., Quiroz,J., Rachin,E., Reeves,K., Reiter,M.A., Reigh,R.,
Rellly,B., Rellly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.D.,
Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Silter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,D.,
Steinle,M., Strong,R., Sutton,A., Sytek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,D., Yoon,L., Yoon,Y.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 171110)
Morley,K.C.
Direct Submission
Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 171110)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23908283.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold, and separated
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GV0K
Center clone name: CH230-466D21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 159707 bases at least Q40
Consensus quality: 161692 bases at least Q30
Consensus quality: 163017 bases at least Q20
Estimated insert size: 161507; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.


```

* (see http://www.bjsgc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
  1 241724: contig of 241724 bp in length.
FEATURES
  source
    1..241724
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-53024"
  misc_feature
    1..2211
    /note="wgs_contig"
BASE COUNT    60533 a 53609 c 56345 g 67996 t 3241 others
ORIGIN
Query Match    59.2%; Score 30.2; DB 2; Length 241724;
Best Local Similarity 74.3%; Pred. No. 0.38;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY
  1 TGCTCCTTTAAGGCTTGATGTCGACACTGTCATGTGACACTTAAG 51
  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 174895 TGCTCCTTTGAAGGCTTGAGTGTGCTGACACTTCTTGTCACCTTAAG 174845

RESULT 6
AL646055/c 193551 bp DNA linear ROD 14-MAR-2002
LOCUS      Mouse DNA sequence from clone RP23-382C18 on chromosome 11,
DEFINITION complete sequence.
ACCESSION  AL646055
VERSION    AL646055.10 GI:19031711
KEYWORDS  HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 193551)
REFERENCE  Kay, M.
AUTHORS   Direct Submission
TITLE     Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL   Cambridgeshire, CB10 1SA. UK. E-mail enquiries:
          humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
          On Feb 28, 2002 this sequence version replaced gi:18476931.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. The following
          abbreviations are used to associate primary accession numbers given
          in the feature table with their source databases: Em., EMBL; Sw.,
          SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
          database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-382C18 is
          from the RPCI-23 Mouse PAC Library
          constructed by the group of Pieter de Jong.
          For further details see http://www.chori.org/bacpac/home.htm
          VECTOR: pBACE3.6
          This sequence is the entire insert of clone RP23-382C18 The true

```

```

left end of clone RP23-70U6 is at 112155 in this sequence.
FEATURES
  source
    1..193551
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="11"
    /clone="RP23-382C18"
    /clone_11b="RPCI-23"
    /complement(86114..86118)
    /note="Sequence from uni-directional primer reads and dGTP
    big dye terminator reads only."
  misc_feature
    160975
    /note="Tandem repeat. Forced join. Assembly confirmed by
    restriction digest data."
BASE COUNT    57068 a 40716 c 39095 g 56672 t
ORIGIN
Query Match    56.9%; Score 29; DB 10; Length 193551;
Best Local Similarity 77.8%; Pred. No. 1.1;
Matches 35; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
OY
  4 TCCTTTAAGGCTTGATGTCGACACTGTCATGTCGACACTTA 48
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 60646 TCAGGATATGGCTTGATGTCGACACTGTCATGTCGACTTA 60602

RESULT 7
AP004394 98286 bp DNA linear HTG 21-MAR-2002
LOCUS      Oryza sativa (japonica cultivar-group) chromosome 2 clone
DEFINITION OJ1118_C06, *** SEQUENCING IN PROGRESS ***.
ACCESSION  AP004394
VERSION    AP004394.1 GI:17402600
KEYWORDS  HTG; PHASE2.
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM  Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharitoidae; Oryzeae; Oryza.
1
REFERENCE  1
AUTHORS   Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE     Oryza sativa Nipponbare(GA3) genomic DNA, chromosome 2, BAC
          clone:OJ1118_C06
          Published Only in Database (2001)
JOURNAL   2 (bases 1 to 98286)
REFERENCE  Sasaki, T., Matsumoto, T. and Yamamoto, K.
AUTHORS   Direct Submission
TITLE     Submitted (05-DEC-2001) Takuji Sasaki, National Institute of
JOURNAL   Agricultural Sciences, Rice Genome Research Program, Kannondai
          2-1-2, Tsukuba, Ibaraki 305-8602, Japan
          (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
          Tel: 81-298-38-7441, Fax: 81-298-38-7468)
COMMENT   The nucleotide sequence of this BAC clone was generated by
          combining Monsanto and RGP-Japan sequencing data.
          NOTE: It currently consists of 1 contigs. Gaps between the contigs
          are represented as runs of N. The order of the pieces is believed
          to be correct as given, however the sizes of the gaps between them
          are based on estimates that have provided by the submitter. This
          sequence will be replaced by the finished sequence as soon as it is
          available and the accession number will be preserved.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
FEATURES
  source
    1..98286
    /organism="Oryza sativa (japonica cultivar-group)"
    /mol_type="genomic DNA"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /chromosome="2"
    /clone="OJ1118_C06"

```


TITLE
JOURNAL
COMMENT

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Huime,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamares,R., Landers,T., Lehocsky,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneu,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L16535

Center Clone name: 113_K_19

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 654 653: contig of 653 bp in length
* 754 753: gap of 100 bp
* 1470 1469: contig of 716 bp in length
* 1570 1569: gap of 100 bp
* 2271 2270: contig of 701 bp in length
* 2371 2370: gap of 100 bp
* 3062 3061: contig of 691 bp in length
* 3162 3161: gap of 100 bp
* 3878 3877: contig of 716 bp in length
* 3978 3977: gap of 100 bp
* 4687 4686: contig of 709 bp in length
* 4787 4786: gap of 100 bp
* 5501 5500: contig of 714 bp in length
* 5601 5600: gap of 100 bp
* 6316 6315: contig of 715 bp in length
* 6416 6415: gap of 100 bp
* 7137 7136: contig of 721 bp in length
* 7237 7236: gap of 100 bp
* 7945 7944: contig of 708 bp in length
* 8045 8044: gap of 100 bp
* 8753 8752: contig of 708 bp in length
* 8853 8852: gap of 100 bp
* 9565 9564: contig of 712 bp in length
* 9665 9664: gap of 100 bp
* 10376 10375: contig of 711 bp in length
* 10476 10475: gap of 100 bp
* 11184 11184: contig of 709 bp in length
* 11285 11284: gap of 100 bp
* 11999 11999: contig of 715 bp in length
* 12000 12099: gap of 100 bp
* 12774 12774: contig of 675 bp in length

12775 12874: gap of 100 bp
12875 13593: contig of 719 bp in length
13594 13693: gap of 100 bp
13694 14400: contig of 707 bp in length
14401 14500: gap of 100 bp
14501 15216: contig of 716 bp in length
15217 15316: gap of 100 bp
15317 16009: contig of 653 bp in length
16010 16109: gap of 100 bp
16110 16806: contig of 697 bp in length
16807 16906: gap of 100 bp
16907 17624: contig of 718 bp in length
17625 17724: gap of 100 bp
17725 18432: contig of 708 bp in length
18433 18532: gap of 100 bp
18533 19230: contig of 698 bp in length
19231 19330: gap of 100 bp
19331 20053: contig of 723 bp in length
20054 20153: gap of 100 bp
20154 20853: contig of 700 bp in length
20854 20953: gap of 100 bp
20954 21665: contig of 712 bp in length
21666 21765: gap of 100 bp
21766 22480: contig of 715 bp in length
22481 22580: gap of 100 bp
22581 23294: contig of 714 bp in length
23295 23394: gap of 100 bp
23395 24094: contig of 700 bp in length
24095 24194: gap of 100 bp
24195 24897: contig of 703 bp in length
24898 24997: gap of 100 bp
24998 25713: contig of 716 bp in length
25714 25813: gap of 100 bp
25814 26529: contig of 716 bp in length
26530 26629: gap of 100 bp
26630 27345: contig of 716 bp in length
27346 27445: gap of 100 bp
27446 28166: contig of 721 bp in length
28167 28266: gap of 100 bp
28267 28984: contig of 718 bp in length
28985 29084: gap of 100 bp
29085 29769: contig of 665 bp in length
29769 29869: gap of 100 bp
29870 30507: contig of 638 bp in length
30508 30607: gap of 100 bp
30608 31327: contig of 720 bp in length
31328 31427: gap of 100 bp
31428 32151: contig of 724 bp in length
32152 32251: gap of 100 bp
32252 32945: contig of 694 bp in length
32946 33045: gap of 100 bp
33046 33760: contig of 715 bp in length
33761 33860: gap of 100 bp
33861 34571: contig of 711 bp in length
34572 34671: gap of 100 bp
34672 35387: contig of 716 bp in length
35388 35487: gap of 100 bp
35488 36185: contig of 698 bp in length
36186 36285: gap of 100 bp
36286 37003: contig of 718 bp in length
37004 37103: gap of 100 bp
37104 37804: contig of 701 bp in length
37805 37904: gap of 100 bp
37905 38619: contig of 715 bp in length
38620 38719: gap of 100 bp
38720 39423: contig of 704 bp in length
39424 39523: gap of 100 bp
39524 40228: contig of 705 bp in length
40229 40328: gap of 100 bp
40329 41042: contig of 714 bp in length
41043 41142: gap of 100 bp
41143 41844: contig of 702 bp in length
41845 41944: gap of 100 bp

*	41945	42661	contig of 717 bp	in length
*	42662	43761	gap of 100 bp	in length
*	42762	43475	contig of 714 bp	in length
*	43476	43576	gap of 100 bp	in length
*	43576	44286	contig of 710 bp	in length
*	44286	44385	gap of 100 bp	in length
*	44386	45098	contig of 713 bp	in length
*	45098	45198	gap of 100 bp	in length
*	45198	45893	contig of 695 bp	in length
*	45894	45993	gap of 100 bp	in length
*	45994	46717	contig of 724 bp	in length
*	46718	46817	gap of 100 bp	in length
*	46818	47192	contig of 679 bp	in length
*	47193	47592	gap of 100 bp	in length
*	47593	48407	contig of 715 bp	in length
*	48408	48407	gap of 100 bp	in length
*	48408	49124	contig of 716 bp	in length
*	49124	49243	gap of 100 bp	in length
*	49244	49240	contig of 697 bp	in length
*	49241	50021	gap of 100 bp	in length
*	50021	50739	contig of 718 bp	in length
*	50739	50839	gap of 100 bp	in length
*	50839	51547	contig of 709 bp	in length
*	51548	51647	gap of 100 bp	in length
*	51648	52344	contig of 697 bp	in length
*	52345	52444	gap of 100 bp	in length
*	52445	53162	contig of 718 bp	in length
*	53163	53262	gap of 100 bp	in length
*	53263	53974	contig of 712 bp	in length
*	53975	54074	gap of 100 bp	in length
*	54075	54786	contig of 706 bp	in length
*	54787	54880	gap of 100 bp	in length
*	54881	55579	contig of 699 bp	in length
*	55580	55679	gap of 100 bp	in length

[illegible][illegible]

Chen, Y., Garand, R., Cappallo, A., Cidre, A., Clazzone, A., Geronzi, M., Geronzi, M., Collins, S., Collamore, A., Egon, A., Focke, P., Hatfield, K., Jax, K., Lazar, J., Dodd, S., Fero, S., Forrester, J., Fitzhugh, W., Gayer, D., Galan, J., Gadya, S., Grimes, S., Goff, S., Goyette, M., Grudun, L., Grand-pierre, N., Hades, R., Hefford, A., Horton, L., Hulme, W., Iliev, I., Johnson, E., Jones, C., Kandi, A., Karatas, A., Kells, C., Lacombe, K., Lamarca, R., Sanders, T., Schorky, J., Levine, R., Ling, M., McClan, C., Macdonald, P., Major, J., Margus, N., Matthews, C., McAtilly, M., Metwin, P., McKenna, K., McPheters, N., Meldrum, J.,

Means, L., Mohova, T., Menega, V., Murphy, T., Naeiro, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Plunkard, P., Pierre, N., Pollard, V.,
 Raymond, C., Rella, K., Rieback, M., Riley, P., Risk, C., Rodov, P.,
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stancu, Thomas, N., Stojanovic, N.,
 Strauss, S., Subramanian, A., Talamas, J., Testayre, S., Thompson, J.,
 Topham, K., Travers, M., Trivett, N., Trullio, J., Vassiliou, H.,
 Viel, R., Vo, A., Wilson, R., Wu, X., Wyman, J., Ye, W., Yoon, J.,
 Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (32-Nov-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 190393)
 Birtu, N., Bastien, V., Bloom, T., Bonusniak, L., Buckhartz, B.,
 Carrara, J., Chang, J., Chazotte, B., Chappel, V., Collin, A.,
 Cook, A., Cooke, P., Dvoretzky, K., Ewark, K., Diaz, S., Eppert, S.,
 Fero, S., Ferreira, P., Fitzpatrick, M., Gabor, C., Gabor, J.,
 Garcia, S., Gird, S., Graham, G., Grand, Pierre, N., Hartz, N.,
 Hauss, B., Horton, L., Hume, W., Jilley, L., Johnston, P., Jones, C.,
 Kamei, A., Karakas, A., Kells, C., Landers, J., Levine, R.,
 Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, T.,
 Matthews, C., McCarthy, M., Melitini, J., Meadows, L., Milosavljevic,
 Menega, V., Murphy, T., Naeiro, J., Nguyen, C., Nguyen, R., Ntutu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Petersen, K., Plunkard, P., Pierre, N., Raymond, C., Rella, K.,
 Rise, C., Rodov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stancu-Thomas, N.,
 Stojanovic, N., Talamas, J., Testayre, S., Thompson, J., Topham, K.,
 Travers, M., Vassiliou, H., Viel, R., Vo, A., Wilson, R., Wu, X.,
 Wyman, J., Yoon, J., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (32-Oct-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 21, 2002 this sequence version replaced gq:42161726.
 All repeats were identified using RepeatMasker:
 Smith, A.P.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WMR
 Web site: <http://www.seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 19L15
 Center clone name: 19L15
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 189671 bases at least q40
 Consensus quality: 189924 bases at least q20
 Consensus quality: 189924 bases at least q20
 Insert size: 172000; average=1p
 Insert size: 190095; sum-of-reads
 quality coverage: 15.5 in q20 bases; average=1p
 quality coverage: 14.0 in q20 bases; sum-of-reads

 * NOTE: this is a "working draft" sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 7293: contig of 7293 bp in length
 * 7294
 * 7303: gap of 100 bp
 * 7304
 * 37157: contig of 29854 bp in length
 * 37158
 * 37257: gap of 100 bp
 * 37258
 * 82905: contig of 45549 bp in length
 * 82907
 * 82908: gap of 45549 bp in length
 * 82909

* 83007 190309: contig of 107303 bp in length.
Location/Qualifiers
1. .190309
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-491H15"
/clone_lib="RPCI-24 Male Mouse BAC"
1. .7203
misc_feature
/note="assembly_fragment"
7304. 37157
misc_feature
/note="assembly_fragment"
37258. 82906
misc_feature
/note="assembly_fragment"
83007. .190309
/note="assembly_fragment"
BASE COUNT 57985 a 36628 c 36057 g 59339 t 300 others
ORIGIN
Query Match 53.3%; Score 27.2; DB 2; Length 190309;
Best Local Similarity 72.9%; Pred. No. 6;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 3 CTTCTTAAGGCTTGAATGCTGCACTGTCATGTACACTTAA 50
DB 120177 CTTCTTCAAGACTTGAAGTCTCATCTGCCAGGTGCACATTCYAA 120130
RESULT 12
AC074174 183036 bp DNA linear HTG 15-JUL-2000
LOCUS AC074174
DEFINITION Mus musculus chromosome 5 clone RP23-123M9 strain C57BL6/J, WORKING
DRAFT SEQUENCE. 21 unordered pieces.
AC074174
AC074174.1 GI:9211284
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 183036)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,O.L., Maduro,V.B.,
Mastralian,S.D., McCloskey,J.C., Ojodu,M.A., Pearson,R.,
Stantibpop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.
NISC Mouse Sequencing Initiative
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 183036)
Green,E.D.
REFERENCE Direct Submission
Submitted (15-JUL-2000) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
COMMENT
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mousehgri.nih.gov
----- Project Information
Center project name: wx
Center clone name: 123M9
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 168781 bases at least Q40
Consensus quality: 172931 bases at least Q30
Consensus quality: 175240 bases at least Q20
Insert size: 20000; agarose-gel
Insert size: 20200; pulse-field-gel
Insert size: 181036; sum-of-contigs

Quality coverage: 4.16x in Q20 bases; agarose-gel
Quality coverage: 4.12x in Q20 bases; pulse-field-gel
Quality coverage: 4.60x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 3005: contig of 3005 bp in length
* 3105: gap of unknown length
* 3106: gap of unknown length
* 3106: contig of 2502 bp in length
* 5607: gap of unknown length
* 5608: gap of 2795 bp in length
* 5708: gap of unknown length
* 8503: gap of 3305 bp in length
* 8602: gap of unknown length
* 11907: gap of 2658 bp in length
* 11908: gap of unknown length
* 12008: gap of 2658 bp in length
* 14665: gap of unknown length
* 14765: gap of 3139 bp in length
* 14766: gap of unknown length
* 17904: gap of 4545 bp in length
* 18004: gap of unknown length
* 18005: gap of 4545 bp in length
* 22549: gap of unknown length
* 22649: gap of unknown length
* 22650: gap of 4536 bp in length
* 27185: gap of unknown length
* 27285: gap of 4713 bp in length
* 31988: gap of unknown length
* 31989: gap of unknown length
* 32098: gap of 4358 bp in length
* 32099: gap of unknown length
* 36457: gap of 4663 bp in length
* 36557: gap of unknown length
* 41220: gap of unknown length
* 41319: gap of 4299 bp in length
* 45618: gap of unknown length
* 45718: gap of 7402 bp in length
* 45719: gap of unknown length
* 45720: gap of unknown length
* 53121: gap of 6055 bp in length
* 53220: gap of unknown length
* 53221: gap of 6055 bp in length
* 59275: gap of unknown length
* 59276: gap of 6065 bp in length
* 59376: gap of unknown length
* 65440: gap of 10958 bp in length
* 65441: gap of unknown length
* 65442: gap of unknown length
* 76498: gap of 16359 bp in length
* 76499: gap of unknown length
* 76599: gap of unknown length
* 92957: gap of 16499 bp in length
* 92958: gap of unknown length
* 93058: gap of unknown length
* 109556: gap of 20074 bp in length
* 109557: gap of unknown length
* 128730: gap of unknown length
* 128731: gap of 20074 bp in length
* 129731: gap of unknown length
* 129830: gap of 17423 bp in length
* 129831: gap of unknown length
* 147353: gap of unknown length
* 147254: gap of unknown length
* 147354: gap of 35683 bp in length.
Location/Qualifiers
1. .183036
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-123M9"
/clone_lib="RPCI mouse BAC library 23"
1. .3005
misc_feature
/note="assembly_fragment"
3105. .5607
misc_feature
/note="assembly_fragment"
5708. 8502
misc_feature
/note="assembly_fragment"
8603. .11907
misc_feature
/note="assembly_fragment"
12008. .14665
misc_feature
/note="assembly_fragment"
14766. .17904

REFERENCE AUTHORS	REFERENCE AUTHORS	TITLE JOURNAL	TITLE JOURNAL
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowls, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Day-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgag, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Nickerson, N., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojodokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalton, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.	Submitted (11-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 4 (bases 1 to 183291)	Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 6 (bases 1 to 183291)	Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 6 (bases 1 to 183291)
REFERENCE AUTHORS	REFERENCE AUTHORS	TITLE JOURNAL	TITLE JOURNAL
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowls, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Day-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgag, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Nickerson, N., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojodokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalton, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.	Submitted (11-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 4 (bases 1 to 183291)	Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 6 (bases 1 to 183291)	Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 6 (bases 1 to 183291)

Query Match

51.0%

Score 26;

DB 3;

Length 183291;

Best Local Similarity

70.0%

Pred. No. 18;

Query Match 51.0%; Score 26; DB 3; Length 183291;
Best Local Similarity 70.0%; Pred. No. 18;

[illegible]

ATTENTION: For the purpose of this project, the following information is required:

1. Project Name: [Project Name]

2. Project Number: [Project Number]

3. Project Location: [Project Location]

4. Project Start Date: [Project Start Date]

5. Project End Date: [Project End Date]

6. Project Manager: [Project Manager]

7. Project Sponsor: [Project Sponsor]

8. Project Budget: [Project Budget]

9. Project Status: [Project Status]

10. Project Description: [Project Description]

11. Project Objectives: [Project Objectives]

12. Project Deliverables: [Project Deliverables]

13. Project Risks: [Project Risks]

14. Project Stakeholders: [Project Stakeholders]

15. Project Communication Plan: [Project Communication Plan]

16. Project Monitoring and Reporting: [Project Monitoring and Reporting]

17. Project Change Management: [Project Change Management]

18. Project Closure: [Project Closure]

19. Project Evaluation: [Project Evaluation]

20. Project Lessons Learned: [Project Lessons Learned]

21. Project Archiving: [Project Archiving]

22. Project Handover: [Project Handover]

23. Project Review: [Project Review]

24. Project Sign-off: [Project Sign-off]

25. Project Completion: [Project Completion]

26. Project Success: [Project Success]

27. Project Impact: [Project Impact]

28. Project Sustainability: [Project Sustainability]

29. Project Innovation: [Project Innovation]

30. Project Collaboration: [Project Collaboration]

31. Project Leadership: [Project Leadership]

32. Project Teamwork: [Project Teamwork]

33. Project Accountability: [Project Accountability]

34. Project Transparency: [Project Transparency]

35. Project Integrity: [Project Integrity]

36. Project Honesty: [Project Honesty]

37. Project Trust: [Project Trust]

38. Project Respect: [Project Respect]

39. Project Compassion: [Project Compassion]

40. Project Kindness: [Project Kindness]

41. Project Generosity: [Project Generosity]

42. Project Gratitude: [Project Gratitude]

43. Project Humility: [Project Humility]

44. Project Patience: [Project Patience]

45. Project Persistence: [Project Persistence]

46. Project Perseverance: [Project Perseverance]

47. Project Determination: [Project Determination]

48. Project Resilience: [Project Resilience]

49. Project Flexibility: [Project Flexibility]

50. Project Adaptability: [Project Adaptability]

51. Project Creativity: [Project Creativity]

52. Project Innovation: [Project Innovation]

53. Project Entrepreneurship: [Project Entrepreneurship]

54. Project Leadership: [Project Leadership]

55. Project Teamwork: [Project Teamwork]

56. Project Accountability: [Project Accountability]

57. Project Transparency: [Project Transparency]

58. Project Integrity: [Project Integrity]

59. Project Honesty: [Project Honesty]

60. Project Trust: [Project Trust]

61. Project Respect: [Project Respect]

62. Project Compassion: [Project Compassion]

63. Project Kindness: [Project Kindness]

64. Project Generosity: [Project Generosity]

65. Project Gratitude: [Project Gratitude]

66. Project Humility: [Project Humility]

67. Project Patience: [Project Patience]

68. Project Persistence: [Project Persistence]

69. Project Perseverance: [Project Perseverance]

70. Project Determination: [Project Determination]

71. Project Resilience: [Project Resilience]

72. Project Flexibility: [Project Flexibility]

73. Project Adaptability: [Project Adaptability]

74. Project Creativity: [Project Creativity]

75. Project Innovation: [Project Innovation]

76. Project Entrepreneurship: [Project Entrepreneurship]

77. Project Leadership: [Project Leadership]

78. Project Teamwork: [Project Teamwork]

79. Project Accountability: [Project Accountability]

80. Project Transparency: [Project Transparency]

81. Project Integrity: [Project Integrity]

82. Project Honesty: [Project Honesty]

83. Project Trust: [Project Trust]

84. Project Respect: [Project Respect]

85. Project Compassion: [Project Compassion]

86. Project Kindness: [Project Kindness]

87. Project Generosity: [Project Generosity]

88. Project Gratitude: [Project Gratitude]

89. Project Humility: [Project Humility]

90. Project Patience: [Project Patience]

91. Project Persistence: [Project Persistence]

92. Project Perseverance: [Project Perseverance]

93. Project Determination: [Project Determination]

94. Project Resilience: [Project Resilience]

95. Project Flexibility: [Project Flexibility]

96. Project Adaptability: [Project Adaptability]

97. Project Creativity: [Project Creativity]

98. Project Innovation: [Project Innovation]

99. Project Entrepreneurship: [Project Entrepreneurship]

100. Project Leadership: [Project Leadership]

101. Project Teamwork: [Project Teamwork]

102. Project Accountability: [Project Accountability]

103. Project Transparency: [Project Transparency]

104. Project Integrity: [Project Integrity]

105. Project Honesty: [Project Honesty]

106. Project Trust: [Project Trust]

107. Project Respect: [Project Respect]

108. Project Compassion: [Project Compassion]

109. Project Kindness: [Project Kindness]

110. Project Generosity: [Project Generosity]

111. Project Gratitude: [Project Gratitude]

112. Project Humility: [Project Humility]

113. Project Patience: [Project Patience]

114. Project Persistence: [Project Persistence]

115. Project Perseverance: [Project Perseverance]

116. Project Determination: [Project Determination]

117. Project Resilience: [Project Resilience]

118. Project Flexibility: [Project Flexibility]

119. Project Adaptability: [Project Adaptability]

120. Project Creativity: [Project Creativity]

121. Project Innovation: [Project Innovation]

122. Project Entrepreneurship: [Project Entrepreneurship]

123. Project Leadership: [Project Leadership]

124. Project Teamwork: [Project Teamwork]

125. Project Accountability: [Project Accountability]

126. Project Transparency: [Project Transparency]

127. Project Integrity: [Project Integrity]

128. Project Honesty: [Project Honesty]

129. Project Trust: [Project Trust]

130. Project Respect: [Project Respect]

131. Project Compassion: [Project Compassion]

132. Project Kindness: [Project Kindness]

133. Project Generosity: [Project Generosity]

134. Project Gratitude: [Project Gratitude]

135. Project Humility: [Project Humility]

136. Project Patience: [Project Patience]

137. Project Persistence: [Project Persistence]

138. Project Perseverance: [Project Perseverance]

139. Project Determination: [Project Determination]

140. Project Resilience: [Project Resilience]

141. Project Flexibility: [Project Flexibility]

142. Project Adaptability: [Project Adaptability]

143. Project Creativity: [Project Creativity]

144. Project Innovation: [Project Innovation]

145. Project Entrepreneurship: [Project Entrepreneurship]

146. Project Leadership: [Project Leadership]

147. Project Teamwork: [Project Teamwork]

148. Project Accountability: [Project Accountability]

149. Project Transparency: [Project Transparency]

150. Project Integrity: [Project Integrity]

151. Project Honesty: [Project Honesty]

152. Project Trust: [Project Trust]

153. Project Respect: [Project Respect]

154. Project Compassion: [Project Compassion]

155. Project Kindness: [Project Kindness]

156. Project Generosity: [Project Generosity]

157. Project Gratitude: [Project Gratitude]

158. Project Humility: [Project Humility]

159. Project Patience: [Project Patience]

160. Project Persistence: [Project Persistence]

161. Project Perseverance: [Project Perseverance]

162. Project Determination: [Project Determination]

163. Project Resilience: [Project Resilience]

164. Project Flexibility: [Project Flexibility]

165. Project Adaptability: [Project Adaptability]

166. Project Creativity: [Project Creativity]

167. Project Innovation: [Project Innovation]

168. Project Entrepreneurship: [Project Entrepreneurship]

169. Project Leadership: [Project Leadership]

170. Project Teamwork: [Project Teamwork]

171. Project Accountability: [Project Accountability]

172. Project Transparency: [Project Transparency]

173. Project Integrity: [Project Integrity]

174. Project Honesty: [Project Honesty]

175. Project Trust: [Project Trust]

176. Project Respect: [Project Respect]

177. Project Compassion: [Project Compassion]

178. Project Kindness: [Project Kindness]

179. Project Generosity: [Project Generosity]

180. Project Gratitude: [Project Gratitude]

181. Project Humility: [Project Humility]

182. Project Patience: [Project Patience]

183. Project Persistence: [Project Persistence]

184. Project Perseverance: [Project Perseverance]

185. Project Determination: [Project Determination]

186. Project Resilience: [Project Resilience]

187. Project Flexibility: [Project Flexibility]

188. Project Adaptability: [Project Adaptability]

189. Project Creativity: [Project Creativity]

190. Project Innovation: [Project Innovation]

191. Project Entrepreneurship: [Project Entrepreneurship]

192. Project Leadership: [Project Leadership]

193. Project Teamwork: [Project Teamwork]

194. Project Accountability: [Project Accountability]

195. Project Transparency: [Project Transparency]

196. Project Integrity: [Project Integrity]

197. Project Honesty: [Project Honesty]

198. Project Trust: [Project Trust]

199. Project Respect: [Project Respect]

200. Project Compassion: [Project Compassion]

201. Project Kindness: [Project Kindness]

202. Project Generosity: [Project Generosity]

203. Project Gratitude: [Project Gratitude]

204. Project Humility: [Project Humility]

205. Project Patience: [Project Patience]

206. Project Persistence: [Project Persistence]

207. Project Perseverance: [Project Perseverance]

208. Project Determination: [Project Determination]

209. Project Resilience: [Project Resilience]

210. Project Flexibility: [Project Flexibility]

211. Project Adaptability: [Project Adaptability]

212. Project Creativity: [Project Creativity]

213. Project Innovation: [Project Innovation]

214. Project Entrepreneurship: [Project Entrepreneurship]

215. Project Leadership: [Project Leadership]

216. Project Teamwork: [Project Teamwork]

217. Project Accountability: [Project Accountability]

218. Project Transparency: [Project Transparency]

219. Project Integrity: [Project Integrity]

220. Project Honesty: [Project Honesty]

221. Project Trust: [Project Trust]

222. Project Respect: [Project Respect]

223. Project Compassion: [Project Compassion]

224. Project Kindness: [Project Kindness]

225. Project Generosity: [Project Generosity]

226. Project Gratitude: [Project Gratitude]

227. Project Humility: [Project Humility]

228. Project Patience: [Project Patience]

229. Project Persistence: [Project Persistence]

230. Project Perseverance: [Project Perseverance]

231. Project Determination: [Project Determination]

232. Project Resilience: [Project Resilience]

233. Project Flexibility: [Project Flexibility]

234. Project Adaptability: [Project Adaptability]

235. Project Creativity: [Project Creativity]

236. Project Innovation: [Project Innovation]

237. Project Entrepreneurship: [Project Entrepreneurship]

238. Project Leadership: [Project Leadership]

239. Project Teamwork: [Project Teamwork]

240. Project Accountability: [Project Accountability]

241. Project Transparency: [Project Transparency]

242. Project Integrity: [Project Integrity]

243. Project Honesty: [Project Honesty]

244. Project Trust: [Project Trust]

245. Project Respect: [Project Respect]

246. Project Compassion: [Project Compassion]

247. Project Kindness: [Project Kindness]

248. Project Generosity: [Project Generosity]

249. Project Gratitude: [Project Gratitude]

250. Project Humility: [Project Humility]

251. Project Patience: [Project Patience]

252. Project Persistence: [Project Persistence]

253. Project Perseverance: [Project Perseverance]

254. Project Determination: [Project Determination]</

```
misc_feature      245022..246206
BASE COUNT      62352 a 43072 c 44760 g 69919 t 35652 others
ORIGIN
Query Match      51.0%: Score 26; DB 2; Length 255755;
Best Local Similarity 70.0%; Pred. No. 18;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY      1 TGCCTCTTTTAAAGGCTTGAAATGCTGCAACTGTGATGTGTACACTTAA 50
      ||||| ||||| ||||| || ||||| ||||| ||||| |||||
Db      190274 TGCCTCTTTTTCCTCCCTGAAGCTAGAACTTTCAGGTGTACTTTAA 190323
```

Search completed: August 20, 2003, 12:34:34
Job time : 267.82 secs

TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Saurin, W., Bernot, A. and Weissenbach, J.	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	Genome Res. 10 (7), 939-949 (2000)	10899143	3 (bases 1 to 902)	Genoscope.	Direct Submission	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .
FEATURES	source	1..902	/organism="Tetraodon nigroviridis"	/mol_type="genomic DNA"	/db_xref="taxon:99883"	/clone="037P04"	/clone_11b="G"	/note="Genoscope sequence ID : COB6037DH02SP1-end : PUC-ori"
BASE COUNT	127 a 156 c 300 g 92 t 227 others							
ORIGIN								
Query Match	13.3%; Score 50; DB 29; Length 902;							
Best Local Similarity	31.6%; Pred. No. 0.0053;							
Matches	99; Conservative 67; Mismatches 147; Indels 0; Gaps 0;							
QY	4 CCCCCGACCCCTATCCTGGCTCCGCCCTTCTCTTCACCCCTCCGAGCCCTAAAGG 63							
Db	881 CCCCCCTCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCSSGSGGG 822							
QY	64 GCGCGCGGGGCCCAAGCGAGGGCGCTGCGCTTACCCCGAGCGGAAGGCCCCAGTGA 123							
Db	821 GGGGGGGGGCCCCCCCCCGSGGGCCCGGGSCCGGGCGGGGGCGGCGGGCGCGG 762							
QY	124 GGTCTTAATGCGGGGTGGCGCTCTCTTGACAGCGCGCGCTTGGGGACAAAGCGGGACG 183							
Db	761 CGSSGSGCGSGGGGGSSGGCGGAGVSSGSSGSGCGMGGGGGMSGMSMGGGGGGS 702							
QY	184 AGAGATAAGGTGACATACACAGACAGATTTGGTGGCGCGGCTGATCTCTCTCCGACA 243							
Db	701 MMGGGVCGMGGMGMVCMCTCTKKMKTTTTRKKKKKTNHMKKTKTMBTKKMKTKKMMK 642							
QY	244 GGAAGCGGGAGCATTTAAAGAACCCATGATTAATTCTTCTCTCGAAGACTCT 303							
Db	641 KKKKKKKKKATATATTKDKMAAVTAMATAAADAAADAAATAKKAARKKAKKATTKGAK 582							
QY	304 TGCGAGAGACAA 316							
Db	581 AKTGTTRKRAAAMA 569							
RESULT 2	CD329933/c	982 bp	mRNA	Linear	EST 28-MAY-2003			
LOCUS	AGNCNCOURT_14164463	NICHD_XGC_Brn1	Xenopus laevis	cdna clone				
DEFINITION	IMAGE:6951625 5', mRNA sequence.							
ACCESSION	CD329933							
VERSION	CD329933.1	GI:31094264						
KEYWORDS	EST.							
SOURCE	Xenopus laevis (African clawed frog)							
ORGANISM	Xenopus laevis							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.							
AUTHORS	1 (bases 1 to 982)							
TITLE	NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC)							

COMMENT	JOURNAL	UNPUBLISHED
CONTACT: DANIELA S. GERHARD, PH.D.	OFFICE OF CANCER GENOMICS	NATIONAL CANCER INSTITUTE / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892	Email: c9apbs-r@mail.nih.gov	Tissue Procurement:
CDNA Library Preparation: Life Technologies, Inc.	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)	DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:	http://image.lnl.gov	plate: LLM14570 row: P column: 24
High quality sequence stop: 297.	Location/Qualifiers	1..982
FEATURES	SOURCE	/organism="Xenopus laevis"
		/mol_type="mRNA"
		/db_xref="taxon:8355"
		/clone="IMAGE:6951625"
		/dev_stage="adult"
		/lab_host="DH10B (phage-resistant)"
		/clone_lib="NICHE_XGC_Brn1"
		/note="Organ: brain; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."
BASE COUNT	172 a	221 c 445 g 89 t 55 others
ORIGIN		
Query Match	13.2%	Score 49.4; DB 14; Length 982;
Best Local Similarity	43.2%	Pred. No. 0.0079;
Matches	95; Conservative	0; Mismatches 125; Indels 0; Gaps 0;
QY	4	CCCCCGCCACCCCTATCTGCGTCCGCGCCCTTCTTCACCCCGCGGACCCCTAAGC 63
DB	543	CCCCCCCCCGCCCCCCCCCTTCCCTCCCGCGCGTCCCGCCCGCCCGCCCGGG 484
QY	64	GGCGCGCGGGCCCAAGCGAGGCGCGTGCCTGACCCCGACCGAAGGGCCCACTTA 123
DB	483	GGCGCGGGGGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 424
QY	124	GGTCTTAATGCGGGTGCCTCTTTTGACAGCGCGCGTTTGGGACAACAGCGGGAC 183
DB	423	GGCGCGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 364
QY	184	AGAGATTAAGTGACATACACAGACGATTTGGTGCAGCG 223
DB	363	NGNGNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 324
RESULT 3		
LOCUS	BX391246	1061 bp mRNA Linear EST 13-MAY-2003
DEFINITION	BX391246 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CSOD1008YNI9 3-PRIME, mRNA sequence.	
ACCESSION	BX391246	
VERSION	BX391246.1	GI:30615383
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 1061)	
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.	
TITLE	Full-length cDNA libraries and normalization	
JOURNAL	Unpublished	
COMMENT	Genoscope - Centre National de Sequencage	
	BP 191 91006 Evry cedex - France	
	Email: seqrel@genoscope.cns.fr, Web : www.genoscope.cns.fr	


```

OY 4 CCCCCGACCCCTCATCTGCTGCTCCGCCCTCTCTCTCCACCCCTCCGAGACCCCTAAAG 63
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1009 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 950
OY 64 GCGGCGGGGGCCCAAGCCGAGCGCTGCGCTGACCCCGAGCGGAAAGGCCAGTCTA 123
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 949 CCTTATGAGVMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 890
OY 124 GGTCTTAATGCGGGTGGCTCTCTCTTGGACAGCGCGCTTGGGAGACAGCGGGGAGC 183
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 889 AAMVAVGGVGGGGVGRMAAAMVAVVAAAAAGAGGGGGGAGAAAAAAGVSSSGRRAR 830
OY 184 AGAGATAAGGTGACATACAGACAGATTTGTGCGCGCGCTGATCTCTCTCCGACA 243
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 829 ARAAMAVGGSSSSSSSSCGMAAAGMGRRGVSSSSCCCSVRCMDVAVACSCMCMCA 770
OY 244 GGAACGCC 252
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 769 GMARCCCB 761

RESULT 6
BX38963 514 bp mRNA linear EST 02-MAY-2003
LOCUS BX38963 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSOD1065YP08 5-PRIME, mRNA sequence.
ACCESSION BX38963
VERSION BX38963.1 GI:30347036
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/InvitrogenCorporation1600
Faraday Avenue Genoscope sequence ID : CSOD1065DH04QP1.
FEATURES
    source
        1..514
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CSOD1065YP08"
            /tissue_type="PLACENTA COT 25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 100 a 70 c 159 g 101 t 84 others
ORIGIN
Query Match 12.1% Score 45.4; DB 13; Length 514;
Best Local Similarity 28.0%; Pred. No. 0.081;
Matches 53; Conservative 62; Mismatches 74; Indels 0; Gaps 0;
OY 7 CCGGACCCCTCATCTGCTCCGCCCTCTCTCCACCCCTCCGAGACCCCTAAAGGCG 66
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 323 CCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 382
OY 67 GCGGCGCCCAAGCGGGGCTGCTGCTGACCCCGAGCGGAGGGGCGCCAGTCACTAGT 126
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 383 SSSSGGGGSGSGSGSSSGSSSGSGSGSGSGSGSGSGSGSGSGSSSSSGSSSG 442
OY 127 CCTAATGCGGGTGGCTCTCTTTGACAGGCGGCTTTGGGAGACAGCGGGGAGCAGA 186

```

```

Db 443 GGGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 502
OY 187 GATAAGTG 195
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 503 GTGGGGGG 511

RESULT 7
CNS01690/c 1201 bp DNA linear GSS 26-JUL-1999
LOCUS CNS01690
DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN15009 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106446
VERSION AL106446.1 GI:5621964
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyroidae; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with The European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billard at CPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
    source
        1..1201
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BACN15009"
            /clone_lib="DrosBAC"
            /plasmid="pBelobAC11"
            /note="end : SP6"
BASE COUNT 252 a 289 c 224 g 107 t 329 others
ORIGIN
Query Match 12.0% Score 45; DB 29; Length 1201;
Best Local Similarity 41.5%; Pred. No. 0.14;
Matches 56; Conservative 32; Mismatches 47; Indels 0; Gaps 0;
OY 13 CCCATCATCTGCTGCTCCGCCCTCTCTCCACCCCTCCGAGACCCCTAAAGGGGGCGG 72
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1138 CCTCTSCKCCBKSGBSGCGCTGTGBCGCCSYCCSCGSGSGCCTGSSGCGCGCCSG 1079
OY 73 GCGCAAGCGGAGGGGCTGCTGACCCCGAGCGGAGGGGCGCCAGTCACTAGTCTAAT 132
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1078 GCSBSHCYGGGKGGGCGSCGCBGTGKGBTKKSGKCCBKBKBSCKGKYGYTT 1019
OY 133 GCGGATGGCGTCTCC 147
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1018 GCKGKKRSTRTKCC 1004

RESULT 8
AL559222 941 bp mRNA linear EST 31-MAY-2003
LOCUS AL559222
DEFINITION AL559222 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CSODJ012YL14 5-PRIME, mRNA sequence.
ACCESSION AL559222
VERSION AL559222.2 GI:31283355
KEYWORDS EST.

```


[illegible]

more information. This sequence belongs to sequence cluster 2460.1. For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=OS00M1460&NCluster=2460.1>. Contact : Feng Liang Email : liang@life-lyon.fr ; <http://fulllength.invitrogen.com/> INVITROGEN CORPORATION 160 Faraday Avenue Genoscope sequence ID : OS00M1460E0P1 Location/Qualifiers
1..969
+/-
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="OS00M146012"
/class_type="PFAT.LIVER"
/dev_stage="fetal"
/clone_id="Homo sapiens PFAT.LIVER"
/note="Organ: liver; Vector: pCMVSPORT1. First strand cDNA was primed with a NotI-sited (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT15 vector. Library was not normalized."

BASH COUNT 244 a 80 g 150 g 257 t 248 others
ORIGIN
query Match 11.7% Score 44; ERF 7; Length 969;
Posit Local Similarity 14.4%; Eval No. 0.24;
Matches 29; Conservative 88; Mismatches 84; Indels 0; Gaps 0;

OY 3 GGCGGCGACCTTCAAGCTGGCTGGCGGCGGTTCCTTGCAACTGGGACCGCTAAAG 42
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 965 SGGCGGCGGCGTCCG 906
Matches 29; Conservative 88; Mismatches 84; Indels 0; Gaps 0;

OY 63 GCGGCGCGGCG 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 SSS 846
Matches 29; Conservative 88; Mismatches 84; Indels 0; Gaps 0;

OY 123 AAGCTGAATGAG 162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 SNNSTSSASISSTSSSSSSSSSSSSAAASTSNINSASSSSNSAATAAASSSNAANS 785
Matches 29; Conservative 88; Mismatches 84; Indels 0; Gaps 0;

OY 183 CAGACATACATGCATATCA 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 SSSSSANSSSSANSSANSSNSA 785

RESULT 10
BX407619/C
Locus
DEFINITION
Accession
Version
Keywords
Organism
Reference
Authors
Title
Journal
Comment

1908 bp mRNA linear EST 15-MAY-2008
BX407619 Homo sapiens PLACENTA Homo sapiens cDNA clone ZS-DE-EY119
5-PRIME, mRNA sequence.
BX407619 GI:40762809
EST.
Homo sapiens (human)
Homo sapiens
Elkarayola; Muzazai; Oforidate; Granatira; Viorietaria; Estelastonia;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1908)
Life Sciences Resource Project, and colleagues.
Full-length cDNA libraries and normalization.
Unpublished
Contact: Genoscope
Genoscope - Centre National de Séquençage
BP 191 91006 Evry-Corbeil - France
Email: seqlab@genoscope.cns.fr; web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4394.1. For
more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=OS00M5200&NCluster=4394.1>. Contact :
Feng Liang Email : liang@life-lyon.fr ; <http://fulllength.invitrogen.com/> INVITROGEN CORPORATION 160
Faraday Avenue Genoscope sequence ID : OS00M5200E0P1.

FEATURES
source
Location/Qualifiers
1. 1000
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="CS0DE05Y119"
/issue_type="PLACENTA"
/note="Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the PCMVSPORT 6 vector. Library was not normalized."

BASE COUNT
ORIGIN
226 a 205 c 283 g 57 t 229 others

Query Match
Best Local Similarity 11.7%; Score 43.8; DB 13; Length 1000;
Pred. No. 0.28;
Matches 24; Conservative 82; Mismatches 43; Indels 0; Gaps 0;

OY 2 GGGCCCCGACCTCATCTGGCTCCGCGCCCTTCTCTCACCCCTCGAGCCCTAAA 61
Db 638 SSSSSSSSSSSSSCCCTTTCSSCCSSCBRTTBBBRYSSSSSSSYSS 579

OY 62 GGGGCGGGGCGCCCAAGCGAGCGCTGACCCGAGCGGAGGCGCCAGTC 121
Db 578 SSSSSSSSSSBMBSSSSSSSSSSSSSSSSSSSSSCSCSSSSSSSGCCSCS 519

OY 122 TAGTCTATATGGGGGCGCTCTCTT 150
Db 518 GRGGCCCCCGSGGCGCGCCSSSY 490

RESULT 11
LOCUS CNS010EM 1009 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03P19 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL098882
AL098882.1 GI:5610493
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1009)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
source
Location/Qualifiers
1. 1009
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03P19"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"
BASE COUNT
ORIGIN
147 a 377 c 178 g 64 t 243 others

Query Match
Best Local Similarity 11.6%; Score 43.6; DB 29; Length 1009;
Pred. No. 0.32;
Matches 66; Conservative 40; Mismatches 77; Indels 0; Gaps 0;

OY 2 GGGCCCCGACCTCATCTGGCTCCGCGCCCTTCTCTCACCCCTCGGAGCCCTAAA 61
Db 751 GSRCCSGCGGGCGGSGTGGGSGCCCGCCCGCCSSCCCGGCGRKCSCVCGCG 692

OY 62 GGGGCGGGGCGCCCAAGCGAGCGCTGACCCCGAGCGGAGGCGCCAGTC 121
Db 691 GGGGCGGGGSGKGMCGCGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 632

OY 122 TAGTCTATATGGGGGCGCTCTCTTACAGCGGCGCTTGGGCAACAGCGGGA 181
Db 631 YACGSSCGSGCTAGTKKTBCTCTTYBKGKTGYBYSTYNTGTGNGGGS 572

OY 182 CGA 184
Db 571 SSR 569

RESULT 12
LOCUS CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL053013
AL053013.1 GI:4934461
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
Location/Qualifiers
1. 925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT
ORIGIN
120 a 61 c 61 g 172 t 511 others

Query Match
Best Local Similarity 11.6%; Score 43.4; DB 29; Length 925;
Pred. No. 0.35;
Matches 21; Conservative 90; Mismatches 48; Indels 0; Gaps 0;

OY 2 GGGCCCCGACCTCATCTGGCTCCGCGCCCTTCTCTCACCCCTCGGAGCCCTAAA 61
Db 571 SSR 569

[illegible][illegible]

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Hordeum.
1 (bases 1 to 567)
Ozturk, N. Z., Michalowski, C. B., Brazille, S., Borchert, C., Palacio, C.
REFERENCE
AUTHORS

Spermatophyta; Magnoliaceae; Triticaceae; Hordeum.

1 (bases 1 to 567)

REFERENCE
AUTHORS

1 (bases 1 to 567)
Ozturk, N. Z., Michalowski, C. B., Brazille, S., Borchert, C., Palacio, C.,
Normand, C., Murphy, C., Kelley, R., Sant, S. A., McLaughlin, H.,
Fredricksen, M. A. and Bohnert, H. J.

REFERENCE	1 (bases 1 to 567)
AUTHORS	Ozluurk, N. Z., Michalowski, C. B., Brazille, S., Borchert, C., Palacio, C.

, Normand, C., Murphy, C., Kelley, R., Sant, S. A., McLaughlin, H.,
Fredricksen, M. A. and Bohnert, H. J.

TITLE Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley

JOURNAL Unpublished (2002)

COMMENT Contact: Mark Fredricksen

Department of Plant Biology

University of Illinois
1001 N. Green St. Urb-

1201 W. Gregory Dr., Urbana
E-1 3173055473

Tel: 41/20554/3
Email: bobhoort1@yahoo.nl

Location (Qualifying)

SOURCE	1	567	LOCATION/QUALITY
--------	---	-----	------------------

source
L. : 20/
/organism="Hordet

```
/organism=Homo sapiens
/mol type="mRNA"
```

/strain="cv toka

/db xref="taxon:"

```

/cjone="HB05B12"

```

```

/tissue_type="Le

```

```
dev_stage="3 we
```

```
/clone_lib="HB"
```

```
/note="6 and 10
```

moist paper (75%)

BASE COUNT	70 a	258 c	154
------------	------	-------	-----

ORIGIN

Query Match	11.48; Scc
-------------	------------

Best Local Similarity 68.6%; Pre

Matches	59;	Conservative	0;
---------	-----	--------------	----

[illegible]

8 CCGACCCCTCAICCTGGCTCCG

155

DU
L J J C C C C C C C C C C C C C C C C

68 CCGGGGCGCAAGCCAGGGCGCT

27

215 GCGGCCCGAGGCCGAGGCCGAG

[illegible]

Search completed: August 20, 2003, 14:28:53
Job time : 1305.11 secs

Job time : 1305.11 secs


```
Db 181 ACAGAGATTAAGTACATACAGAGCAGATTGGTGGCGCGCGTGAATCTCTCCCG 240
QY 241 ACAGAAACGGGAGCTTTTAAAGACCCATGATTTACTTTATCTTCTCTGAAAGCT 300
Db 241 ACAGAAACGGGAGCTTTTAAAGACCCATGATTTACTTTATCTTCTCTGAAAGCT 300
QY 301 TCTTGGGAGAGACAAAGATGTCTCCCTGCTTAAGACACAAGGCGCACAAAGAGGCT 360
Db 301 TCTTGGGAGAGACAAAGATGTCTCCCTGCTTAAGACACAAGGCGCACAAAGAGGCT 360
QY 361 CTGCACAGGCGAGCG 375
Db 361 CTGCACAGGCGAGCG 375
```

```
RESULT 2
US-09-761-466-3
; Sequence 3, Application US/09761466
; Patent No. US20020022259A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Ike W.
; APPLICANT: Izumo, Seigo
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; FILE REFERENCE: 01948/069002
; CURRENT APPLICATION NUMBER: US/09/761,466
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1072
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-761-466-3
```

```
Query Match 100.0%; Score 375; DB 9; Length 1072;
Best Local Similarity 100.0%; Pred. No. 1.8e-112;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AGGCCCCCGGACCCATCTGGGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTAA 60
Db 1 AGGCCCCCGGACCCATCTGGGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTAA 60
QY 61 AGGCGCGGCGGGGCCCAAGCGAGGCGCTGCCCTTGACCCCGAGCGGAAGGCGCCAGT 120
Db 61 AGGCGCGGCGGGGCCCAAGCGAGGCGCTGCCCTTGACCCCGAGCGGAAGGCGCCAGT 120
QY 121 CTAGTCTCTAATGCGGGTGGCGTCTCTTTGACAGCGGGGTTGGGGACAACAGCGGG 180
Db 121 CTAGTCTCTAATGCGGGTGGCGTCTCTTTGACAGCGGGGTTGGGGACAACAGCGGG 180
QY 181 ACAGAGATTAAGTGAATACACAGAGAGATTTGGTCCGCGCTGATCTCTCTCCG 240
Db 181 ACAGAGATTAAGTGAATACACAGAGAGATTTGGTCCGCGCTGATCTCTCTCCG 240
QY 241 ACAGAAACGGGAGCTTTTAAAGACCCATGATTTACTTTATCTTCTCTGAAAGCT 300
Db 241 ACAGAAACGGGAGCTTTTAAAGACCCATGATTTACTTTATCTTCTCTGAAAGCT 300
QY 301 TCTTGGGAGAGACAAAGATGTCTCCCTGCTTAAGACACAAGGCGCACAAAGAGGCT 360
Db 301 TCTTGGGAGAGACAAAGATGTCTCCCTGCTTAAGACACAAGGCGCACAAAGAGGCT 360
QY 361 CTGCACAGGCGAGCG 375
Db 361 CTGCACAGGCGAGCG 375
```

RESULT 3
US-09-761-466-4

```
; Sequence 4, Application US/09761466
; Patent No. US20020022259A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Ike W.
; APPLICANT: Izumo, Seigo
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; FILE REFERENCE: 01948/069002
; CURRENT APPLICATION NUMBER: US/09/761,466
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(7838)
; OTHER INFORMATION: n = A,T,C or G
US-09-761-466-4
```

```
Query Match 100.0%; Score 375; DB 9; Length 7838;
Best Local Similarity 100.0%; Pred. No. 3.5e-112;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AGGCCCCCGGACCCATCTGGGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTAA 60
Db 3588 AGGCCCCCGGACCCATCTGGGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTAA 3647
QY 61 AGGCGCGGCGGGGCCCAAGCGAGGCGCTGCCCTTGACCCCGAGCGGAAGGCGCCAGT 120
Db 3648 AGGCGCGGCGGGGCCCAAGCGAGGCGCTGCCCTTGACCCCGAGCGGAAGGCGCCAGT 3707
QY 121 CTAGTCTCTAATGCGGGTGGCGTCTCTTTGACAGGCGGGGTTGGGGACAACAGCGGG 180
Db 3708 CTAGTCTCTAATGCGGGTGGCGTCTCTTTGACAGGCGGGGTTGGGGACAACAGCGGG 3767
QY 181 ACAGAGATTAAGTGAATACACAGAGAGATTTGGTCCGCGCTGATCTCTCTCCG 240
Db 3768 ACAGAGATTAAGTGAATACACAGAGAGATTTGGTCCGCGCTGATCTCTCTCCG 3827
QY 241 ACAGAAACGGGAGCTTTTAAAGACCCATGATTTACTTTATCTTCTCTGAAAGCT 300
Db 3828 ACAGAAACGGGAGCTTTTAAAGACCCATGATTTACTTTATCTTCTCTGAAAGCT 3887
QY 301 TCTTGGGAGAGACAAAGATGTCTCCCTGCTTAAGACACAAGGCGCACAAAGAGGCT 360
Db 3888 TCTTGGGAGAGACAAAGATGTCTCCCTGCTTAAGACACAAGGCGCACAAAGAGGCT 3947
QY 361 CTGCACAGGCGAGCG 375
Db 3948 CTGCACAGGCGAGCG 3962
```

```
RESULT 4
US-10-237-496-101/C
; Sequence 101, Application US/10237496
; Publication No. US20030138896A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
```



```
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 101
; LENGTH: 1111
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-243-402-101
```

```
Query Match          9.8%; Score 36.6; DB 12; Length 1111;
Best Local Similarity 49.7%; Pred. No. 0.093;
Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```
QY 53 ACCCCCTAAGGGGGGGGGGGGGCCCAAGCCGAGGGGGCTGGGCGCTGACCCCGAGGGGAAG 112
DB 466 AGCAGCGAAGGGCCCGGACCTCGCACCCAGCTGGGGCCCTGGCAGCGTCCGAGGCCAGC 407
QY 113 GCCCCTAAGTCTCTATATGCGGGGTGGCTCTCTTGTGACAGCGCGGCTTTGGGGAACA 172
DB 406 AGCCAGTCCCGGCGATATATCCCGAGCGGGGTGATGAGAGGGGCAAGGTTGGGGAACG 347
QY 173 CAGCGGGGACGAGATAGGTGACATACAGAGCAGATTGGTGGCGCGCTATACTC 232
DB 346 CCCCCGCAAAAGCATTGGGGGTACCTGGCGGACGCAATTGACTTCCCTCAGAGTCCCGC 287
QY 233 CTCTCCG 239
DB 286 CTCTCGC 280
```

```
RESULT 11
US-10-243-431-101/c
; Sequence 101, Application US/10243431
; Publication No. US20030138903A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Baton, Dan
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C37
; CURRENT APPLICATION NUMBER: US/10/243,431
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
```

```
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 101
; LENGTH: 1111
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-243-431-101
```

```
Query Match          9.8%; Score 36.6; DB 12; Length 1111;
Best Local Similarity 49.7%; Pred. No. 0.093;
Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```
QY 53 ACCCCCTAAGGGGGGGGGGGGGCCCAAGCCGAGGGGGCTGGGCGCTGACCCCGAGCGGAAG 112
DB 466 AGCAGCGAAGGGCCCGGACCTCGCACCCAGCTGGGGCCCTGGCAGCGTCCGAGGCCAGC 407
QY 113 GCCCCTAAGTCTCTATATGCGGGGTGGCTCTCTTGTGACAGCGGGGCTTTGGGGAACA 172
DB 406 AGCCAGTCCCGGCGATATATCCCGAGCGGGGTGATGAGAGGGGCAAGGTTGGGGAACG 347
QY 173 CAGCGGGGACGAGATAGGTGACATACAGAGCAGATTGGTGGCGCGCTATACTC 232
DB 346 CCCCCGCAAAAGCATTGGGGGTACCTGGCGGACGCAATTGACTTCCCTCAGAGTCCCGC 287
QY 233 CTCTCCG 239
DB 286 CTCTCGC 280
```

```
RESULT 12
US-10-245-164-101/c
; Sequence 101, Application US/10245164
; Publication No. US20030138904A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Baton, Dan
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C76
; CURRENT APPLICATION NUMBER: US/10/245,164
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
```



```
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 101
; LENGTH: 1111
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-107-101

Query Match
Best Local Similarity 49.8%; Score 36.6; DB 14; Length 1111;
Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 53 ACCCTTAAGGGGCGGGCCCAAGCCGAGGGCGCTGACCCCGAGCGGAAG 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 466 ACCACGGAAGGGCCGAGCCCTGCGACCCAGCTGGGCCCTGCGACGCCAGC 407
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 GCCCAGTCTAGTCCCTAATGCGGGTGCGCTCTCTTTGACAGCGCGCTTTGGGACAA 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 406 AGCCCACTCCCGCCATATCCCGGAGCGGGGTCTATGAGAGAGGCGAGGGTTGGGAAAG 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 173 CAGCGGGGACGAGATTAAGTGACATACAGACGAGATTGGTGGCGCGCTGATATCTC 232
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 346 CCCCCGCAAAAGCCATTGGCGGTACCTGGCGAGCGATTGACTTCCCTCAGAGATCCCGC 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 233 CTCTCCC 239
   | | | | |
DB 286 CTCTCGC 280

RESULT 15
US-10-245-143-101/c
; Sequence 101, Application US/10245143
; Publication NO. US20030068780A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gunney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C90
; CURRENT APPLICATION NUMBER: US/10/245,143
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
```

```
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 101
; LENGTH: 1111
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-143-101

Query Match
Best Local Similarity 49.7%; Score 36.6; DB 14; Length 1111;
Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 53 ACCCTTAAGGGGCGGGCCCAAGCCGAGGGCGCTGACCCCGAGCGGAAG 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 466 AGCAGGGAAGGGCCGAGCCCTGCGACCCAGCTGGGCCCTGCGACGCCAGC 407
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 GCCCAGTCTAGTCCCTAATGCGGGTGCGCTCTCTTTGACAGCGCGCTTTGGGACAA 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 406 AGCCCACTCCCGCCATATCCCGGAGCGGGGTCTATGAGAGAGGCGAGGGTTGGGAAAG 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 173 CAGCGGGGACGAGATTAAGTGACATACAGACGAGATTGGTGGCGCGCTGATATCTC 232
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 346 CCCCCGCAAAAGCCATTGGCGGTACCTGGCGAGCGATTGACTTCCCTCAGAGATCCCGC 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 233 CTCTCCC 239
   | | | | |
DB 286 CTCTCGC 280
```

Search completed: August 20, 2003, 18:13:15
Job time : 348.431 secs


```
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aildong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/09/620,312D
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: PL_FL_genes Version 1.0
SEQ ID NO 849
LENGTH: 1281
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (416)..(1033)
US-09-620-312D-849
```

Query Match 9.3%: Score 35; DB 4; Length 1281;

Best Local Similarity 56.5%: Pred. No. 0.32; 50; Indels 0; Gaps 0;

Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 5 CCCCCGACCCCTGCTGCTCCGCCCTTCTCTCCACCCCTCCGACCCCTAAAGG 64

DB 295 CTCCTTCCCTCCCTCCGCCGCCGACGCCGCCGACGCCCTCTCAAGC 354

QY 65 GCGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 119

DB 355 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 409

RESULT 3

US-09-819-989-3

Sequence 3, Application US/09819989

Patent No. 6482629

GENERAL INFORMATION:

APPLICANT: WEI, Ming-Hui et al.

TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC

TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES

TITLE OF INVENTION: THEREOF

FILE REFERENCE: C1001200

CURRENT APPLICATION NUMBER: US/09/819,989

CURRENT FILING DATE: 2001-03-29

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 19650

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(19650)

OTHER INFORMATION: n = A,T,C or G

Query Match 9.0%: Score 33.6; DB 4; Length 19650;

Best Local Similarity 52.1%: Pred. No. 2.8;

Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
QY 1 AGGCCCCCGACCCCTGCTGCTCCGCCCTTCTCTCCACCCCTCCGAGCCCTAA 60
DB 1885 AGAGCCCTGCGGCGGCTCTCCCTCTCCGCCGCCGACCCCTCCGAGCGGCGGCTCCGC 1944
QY 61 AGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 1945 GTGGGCGGCTGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2004
QY 121 CTAGCTCTAATGCGGCTGCGTC 144
DB 2005 GCGGCGCTGATGCTGCTGACGGC 2028
```

RESULT 4

US-09-620-312D-358

Sequence 358, Application US/09620312D

Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyan

APPLICANT: Chen, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aildong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yundong

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: John Tillinghast

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. 6569662el Nucleic Acids and

TITLE OF INVENTION: Polypeptides

FILE REFERENCE: 784CIP2B

CURRENT APPLICATION NUMBER: US/09/620,312D

CURRENT FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/488,725

PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1105

SOFTWARE: PL_FL_genes Version 1.0

SEQ ID NO 358

LENGTH: 2780

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (219)..(2045)

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(2780)

OTHER INFORMATION: n = a,t,c or g

US-09-620-312D-358

Query Match 8.9%: Score 33.2; DB 4; Length 2780;

Best Local Similarity 53.0%: Pred. No. 1.6;

Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 52 GACCCCTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 111

DB 106 GCGGCGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 165

QY 112 GCGCCACTGCTAGCTCTCAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 171

DB 166 ATCCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 225

QY 172 ACAGCGGGAGCGAG 185


```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-684-5600
TELEFAX: 703-836-5288
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Caulobacter crescentus
STRAIN: CB 15
US-08-194-290-6

Query Match
Best Local Similarity 51.7%; Pred. No. 2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 43 ACCCTCCCGAGCCCTTAAGGAGGCGGCGGCCCAAGCCGAGGAGCGCTGCGCTGACCCC 102
DB 1115 ACCGCCCTGAAACACCAACACAGCGGCGGCTCAACCGTCACCGCGCGGCTGGCCAG 1174
QY 103 GAGCGGAAGGCGCCCACTAGTCTTAATGCGGGTGGCGCTCTTGTGACAGCGCGCT 162
DB 1175 AACCTGACCGCCACAGACCGCGGCTCAAGCCGCGAACAACGTCGCGAGCGGCGGCC 1234
QY 163 TTGGGGACACAGCGGCGGAGAGAG 187
DB 1235 AACGTACCGCTCGCTCGACGCGCG 1259

RESULT 8
US-08-614-377A-6
Sequence 6, Application US/08614377A
Patent No. 5976864
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: No. 5976864ellini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF
HETEROLOGOUS POLYPEPTIDES FROM CAULOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson PC
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,377A
FILING DATE: 12-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/194,290
FILING DATE: 09-FEB-1994
CLASSIFICATION: 435
APPLICATION NUMBER: US 07/895,367
FILING DATE: 09-JUNE-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34053
REFERENCE/DOCKET NUMBER: 08106/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
```

```
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Caulobacter crescentus
STRAIN: CB 15
US-08-614-377A-6

Query Match
Best Local Similarity 51.7%; Pred. No. 2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 43 ACCCTCCCGAGCCCTTAAGGAGGCGGCGGCCCAAGCCGAGGAGCGCTGCGCTGACCCC 102
DB 1115 ACCGCCCTGAAACACCAACACAGCGGCGGCTCAACCGTCACCGCGCGGCTGGCCAG 1174
QY 103 GAGCGGAAGGCGCCCACTAGTCTTAATGCGGGTGGCGCTCTTGTGACAGCGCGCT 162
DB 1175 AACCTGACCGCCACAGACCGCGGCTCAAGCCGCGAACAACGTCGCGTGAAGGCGGCC 1234
QY 163 TTGGGGACACAGCGGCGGAGAGAG 187
DB 1235 AACGTACCGCTCGCTCGACGCGCG 1259
```

```
RESULT 9
US-09-142-648B-6
Sequence 6, Application US/09142648B
Patent No. 6210948
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: No. 6210948ellini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CA
FILE REFERENCE: 08106/002002
CURRENT APPLICATION NUMBER: US/09/142,648B
CURRENT FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: PCT/CA97/00167
PRIOR FILING DATE: 1997-03-10
PRIOR APPLICATION NUMBER: US 07/614,377
PRIOR FILING DATE: 1996-03-12
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 3300
TYPE: DNA
ORGANISM: Caulobacter crescentus
FEATURE:
NAME/KEY: CDS
LOCATION: (101)...(3178)
US-09-142-648B-6

Query Match
Best Local Similarity 51.7%; Pred. No. 2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 43 ACCCTCCCGAGCCCTTAAGGAGGCGGCGGCCCAAGCCGAGGAGCGCTGCGCTGACCCC 102
DB 1115 ACCGCCCTGAAACACCAACACAGCGGCGGCTCAACCGTCACCGCGCGGCTGGCCAG 1174
QY 103 GAGCGGAAGGCGCCCACTAGTCTTAATGCGGGTGGCGCTCTTGTGACAGCGCGCT 162
DB 1175 AACCTGACCGCCACAGACCGCGGCTCAAGCCGCGAACAACGTCGCGTGAAGGCGGCC 1234
QY 163 TTGGGGACACAGCGGCGGAGAGAG 187
DB 1235 AACGTACCGCTCGCTCGACGCGCG 1259
```


MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,434A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weseman, James C.
REGISTRATION NUMBER: 30,507
REFERENCE/DOCKET NUMBER: P00060US0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 699-3604
TELEFAX: 619-236-1048
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2875 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-458-434A-4

Query Match 8.6%; Score 32.4; DB 3; Length 2875;
Best Local Similarity 50.6%; Pred. No. 2.9; Mismatches 0; Gaps 0;
Matches 78; Conservative 0; Indels 76; Gaps 0;

Y 5 CCCCCGACCCCTATCTGCTCCGCCCTTCTCTCCACCCCTCCGAGCCCTAAAGG 64
DB 1791 CCCCCGACCCCTATCTGCTCCGCCCTTCTCTCCACCCCTCCGAGCCCTAAAGG 1850
Y 65 GCGCGGGGCGCCCAAGCGAGGCGCTGCGCTGACCCCGGAGGCGCCCAAGTCTAG 124
DB 1851 CCGCGGGGCGGATCCCGCGGCTGCTCCGAGGGGAGGTTCGCCACGCGCGGA 1910
Y 125 GTCTTAATGCGGGTGCGTCTCTTTGACAGCG 158
DB 1911 GGGAGCCGCGAGCGGCGTCTCTTTAAAGCG 1944

RESULT 14
US-08-458-434A-6
Sequence 6, Application US/08458434A
Patent No. 6083690
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Stephen E.
APPLICANT: Mundy M.D., Gregory R.
APPLICANT: Gosh-Choudhury Ph.D., Nandini
APPLICANT: Feng Ph.D., Jian Q.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: James C. Weseman, Esq.
STREET: 401 B. Street, Suite 1700
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,434A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weseman, James C.
REGISTRATION NUMBER: 30,507
REFERENCE/DOCKET NUMBER: P00060US0

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 699-3604
TELEFAX: 619-236-1048
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1514 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-458-434A-6

Query Match 8.6%; Score 32.4; DB 3; Length 1514;
Best Local Similarity 50.6%; Pred. No. 5.9;
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Y 5 CCCCCGACCCCTATCTGCTCCGCCCTTCTCTCCACCCCTCCGAGCCCTAAAGG 64
DB 1791 CCCCCGACCCCTATCTGCTCCGCCCTTCTCTCCACCCCTCCGAGCCCTAAAGG 1850
Y 65 GCGCGGGGCGCCCAAGCGAGGCGCTGCGCTGACCCCGGAGGCGCCCAAGTCTAG 124
DB 1851 CCGCGGGGCGGATCCCGCGGCTGCTCCGAGGGGAGGTTCGCCACGCGCGGA 1910
Y 125 GTCTTAATGCGGGTGCGTCTCTTTGACAGCG 158
DB 1911 GGGAGCCGCGAGCGGCGTCTCTTTAAAGCG 1944

RESULT 15
US-09-252-991A-12955
Sequence 12955, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 12955
LENGTH: 846
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12955

Query Match 8.6%; Score 32.2; DB 4; Length 846;
Best Local Similarity 61.2%; Pred. No. 1.9;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Y 2 GGGCCCCGACCCCTATCTGCTCCGCCCTTCTCTCCACCCCTCCGAGCCCTAA 61
DB 14 GGACACCGGGCGCATGTTATTTCACAGCGCTTCTGCGCTTCTCGGAGCCAGAAC 73
Y 62 GGGCGCGCGGCGCCCAAGCGAGG 86
DB 74 AGGTGGGCGGCGGCGAGCGGCGG 98

Search completed: August 20, 2003, 14:32:38
Job time: 53.7146 secs

0.000000 version 5.1.6
0.000000 (0) 0.000000 0.000000 0.000000

Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Maximum number of iterations: 1000000
Maximum number of iterations: 1000000
Maximum number of iterations: 1000000

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the results being printed, and is derived by analysis of the total score distribution.

SUMMARY

Prod. No.	Score	Match	Length	IB	IL	Location
1	135	135	135	22	AA53958	Human chromosome 10p11.2
2	135	135	135	22	AA53958	Human chromosome 10p11.2
3	135	135	135	22	AA53958	Human chromosome 10p11.2
4	135	135	135	22	AA53958	Human chromosome 10p11.2
5	135	135	135	22	AA53958	Human chromosome 10p11.2
6	135	135	135	22	AA53958	Human chromosome 10p11.2
7	135	135	135	22	AA53958	Human chromosome 10p11.2
8	135	135	135	22	AA53958	Human chromosome 10p11.2
9	135	135	135	22	AA53958	Human chromosome 10p11.2
10	135	135	135	22	AA53958	Human chromosome 10p11.2

ATP-GENES

Prod. No.	Score	Match	Length	IB	IL	Location
1	135	135	135	22	AA53958	Human chromosome 10p11.2
2	135	135	135	22	AA53958	Human chromosome 10p11.2
3	135	135	135	22	AA53958	Human chromosome 10p11.2
4	135	135	135	22	AA53958	Human chromosome 10p11.2
5	135	135	135	22	AA53958	Human chromosome 10p11.2
6	135	135	135	22	AA53958	Human chromosome 10p11.2
7	135	135	135	22	AA53958	Human chromosome 10p11.2
8	135	135	135	22	AA53958	Human chromosome 10p11.2
9	135	135	135	22	AA53958	Human chromosome 10p11.2
10	135	135	135	22	AA53958	Human chromosome 10p11.2

PT induction, e.g. for optimizing cardiomyocyte induction
 XX
 PS Claim 1; Fig 5A; 66pp; English.
 XX
 CC The sequence represents the coding sequence of cardiac enhancer
 CC mcsx/Nkx2.5 homology domain A1. The nucleic acid is useful for
 CC specifically expressing a gene in a cardiac cell, as an earlier marker
 CC of cardiomyocyte induction, e.g. for optimizing cardiomyocyte induction.
 CC Genes expressed in the cardiac cell-specific manner are useful for the
 CC targeted expression of genes encoding therapeutic proteins for the
 CC treatment of damaged heart tissue. Cardiac specific enhancer elements may
 CC be used for gene therapy.
 CC
 XX Sequence 375 BP; 82 A; 120 C; 108 G; 65 T; 0 other:
 SQ
 Query Match 100.0%; Score 375; DB 22; Length 375;
 Best Local Similarity 100.0%; Pred. No. 1.2e-100;
 Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AGGCCCCCGCACCCCTCATCTGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTTAA 60
 DB 1 AGGCCCCCGCACCCCTCATCTGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTTAA 60
 OY 61 AGGGCGGGGGGGCCCAAGCCGAGGGCGCTGCGCCTGACCCCGAGCGGAAGGCCCCAGT 120
 DB 61 AGGGCGGGGGGGCCCAAGCCGAGGGCGCTGCGCCTGACCCCGAGCGGAAGGCCCCAGT 120
 OY 121 CTAGTCTCTAATGCGGGTGGCGCTCTCTTGTGACAGCGGGCGTTGGGGGACACAGCGGGG 180
 DB 121 CTAGTCTCTAATGCGGGTGGCGCTCTCTTGTGACAGCGGGCGTTGGGGGACACAGCGGGG 180
 OY 181 ACAGAGATAGGTGACATACAGACAGATTGGTGGCGGCGCTGATCTCTCTCCG 240
 DB 181 ACAGAGATAGGTGACATACAGACAGATTGGTGGCGGCGCTGATCTCTCTCCG 240
 OY 241 ACAGGAACCGCGAGCTATTAAAAAGCCCTATTCATTATCTTTCTGGAAGCT 300
 DB 241 ACAGGAACCGCGAGCTATTAAAAAGCCCTATTCATTATCTTTCTGGAAGCT 300
 OY 301 TCTTGGGAGAGACAAAGAATGTTCCCTGCTTAAAGACAAAGGCCACACAGGAGGT 360
 DB 301 TCTTGGGAGAGACAAAGAATGTTCCCTGCTTAAAGACAAAGGCCACACAGGAGGT 360
 OY 361 CTGCACAGCGGACGC 375
 DB 361 CTGCACAGCGGACGC 375
 RESULT 2
 AAS09960 standard; DNA: 1072 BP.
 XX
 AC AAS09960;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE DNA encoding mcsx/Nkx2.5 homology domains A1 + A2, and intervening DNA.
 XX
 KW mcsx/Nkx2.5 homology domain; cardiac enhancer; cardiac cell;
 KW cardiomyocyte induction; therapeutic; heart tissue; gene therapy;
 KW mouse; ds.
 XX
 OS Mus musculus.
 XX
 PN WO200151006-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 16-JAN-2001; 2001WO-US01511.
 XX
 PR 14-JAN-2000; 2000US-0176419.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX
 XX Lee IW, Izumo S;
 XX
 DR WPI: 2001-451809/48.
 XX
 PT New cardiac specific cell enhancer elements, useful for specifically
 PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
 PT induction, e.g. for optimizing cardiomyocyte induction
 XX
 PS Claim 1; Fig 5B; 66pp; English.
 XX
 CC The sequence represents the coding sequence of cardiac enhancer
 CC mcsx/Nkx2.5 homology domains A1 + A2 and intervening sequence. The
 CC nucleic acid is useful for specifically expressing a gene in a cardiac
 CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
 CC optimizing cardiomyocyte induction. Genes expressed in the cardiac cell-
 CC specific manner are useful for the targeted expression of genes encoding
 CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
 CC specific enhancer elements may be used for gene therapy.
 CC
 XX Sequence 1072 BP; 273 A; 290 C; 297 G; 212 T; 0 other:
 SQ
 Query Match 100.0%; Score 375; DB 22; Length 1072;
 Best Local Similarity 100.0%; Pred. No. 1.7e-100;
 Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AGGCCCCCGCACCCCTCATCTGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTTAA 60
 DB 1 AGGCCCCCGCACCCCTCATCTGCTCCGCGCCCTTCTCTCCACCCCTCCGAGCCCTTAA 60
 OY 61 AGGGCGGGGGGGCCCAAGCCGAGGGCGCTGCGCCTGACCCCGAGCGGAAGGCCCCAGT 120
 DB 61 AGGGCGGGGGGGCCCAAGCCGAGGGCGCTGCGCCTGACCCCGAGCGGAAGGCCCCAGT 120
 OY 121 CTAGTCTCTAATGCGGGTGGCGCTCTCTTGTGACAGCGGGCGTTGGGGGACACAGCGGGG 180
 DB 121 CTAGTCTCTAATGCGGGTGGCGCTCTCTTGTGACAGCGGGCGTTGGGGGACACAGCGGGG 180
 OY 181 ACAGAGATAGGTGACATACAGACAGATTGGTGGCGGCGCTGATCTCTCTCCG 240
 DB 181 ACAGAGATAGGTGACATACAGACAGATTGGTGGCGGCGCTGATCTCTCTCCG 240
 OY 241 ACAGGAACCGCGAGCTATTAAAAAGCCCTATTCATTATCTTTCTGGAAGCT 300
 DB 241 ACAGGAACCGCGAGCTATTAAAAAGCCCTATTCATTATCTTTCTGGAAGCT 300
 OY 301 TCTTGGGAGAGACAAAGAATGTTCCCTGCTTAAAGACACAGGCCACACAGGAGGT 360
 DB 301 TCTTGGGAGAGACAAAGAATGTTCCCTGCTTAAAGACACAGGCCACACAGGAGGT 360
 OY 361 CTGCACAGCGGACGC 375
 DB 361 CTGCACAGCGGACGC 375
 RESULT 3
 AAS09961 standard; DNA: 7836 BP.
 XX
 AC AAS09961;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Genomic DNA #1 encoding human Csx/Nkx2.5.
 XX
 KW Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
 KW therapeutic; heart tissue; gene therapy; human; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200151006-A2.
 XX
 PD 19-JUL-2001.

XX The present sequence is that of the human ZFP36 (homologous to zfp-36 in
CC mouse) gene that encodes the ZFP36 protein of the invention (see #
CC AB080316). The specification describes a newly isolated polynucleotide
CC comprising a sequence which is a polymorphic variant (PV) of a reference
CC sequence for the ZFP36 gene or its fragment and its encoded protein. The
CC ZFP36 polynucleotides and polypeptides have antirheumatic,
CC immunosuppressive and antiarthritic activities. The ZFP36 polypeptide is
CC useful for screening drugs targeting the ZFP36 polypeptide. ZFP36
CC isogenes or haplotype pairs are useful for improving the efficiency and
CC reliability of the discovery and development of drugs for treating
CC diseases associated with ZFP36 activity, e.g., autoimmune diseases such
CC as rheumatoid arthritis. Haplotyping the ZFP36 gene in an individual
CC gives useful information for validating ZFP36 as a candidate target for
CC creating a specific condition predicted to be associated with ZFP36
CC activity. Genotyping the ZFP36 gene of an individual can give information
CC used for developing diagnostic tests and therapeutic treatments. The
CC isolated polynucleotide is useful in studying the expression and function
CC of ZFP36 and in drug screening. Antibodies specific for the ZFP36 protein
CC are useful in many diagnostic and prognostic formats and therapeutic
CC methods. A recombinant non-human organism transformed with the ZFP36 gene
CC is useful in studying expression of the ZFP36 isogenes in vivo, for in
CC vivo drug screening and testing. Allele-specific oligonucleotides (ASO)
CC are useful as probes and primers and for assaying a polymorphism in the
CC target region.

XX
SQ Sequence 3889 BP; 648 A; 1304 C; 1078 G; 850 T; 9 other:

Query Match 11.7%; Score 43.8; DB 24; Length 3889;
Best Local Similarity 59.1%; Pred. No. 0.012;
Matches 75; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 2 GGCCCCCGACCCATCTGCTCCGCCCTTCCTCCACCTCCGAGCCCTAA 61
DB 301 GGGCAGGCGTCCCGCATCCGACACCCACCTTCCCGCAGCATTCCTCCGCTGCTAC 360
OY 62 GGGCGCGGCGGCGCCAGCGGCGCTCGGCTGACCCCGAGCGGAGGCCCGCAGTC 121
DB 361 GCGTGTCCACCGGCGCAAGCTCAGGCGGCTCTCCAGCGCGGCGGAGGAAAGCAGTCC 420
OY 122 TAGGTCC 128
DB 421 CAGGGCC 427

RESULT 5
ABK91535
ID ABR91535 standard; DNA; 7552 BP.

XX
AC ABR91535;

XX
DT 14-AUG-2002 (first entry)

DE Modified HIV protein-encoding plasmid DNA #87.

XX
KW HIV; human immunodeficiency virus; gene; ds: circular; cyclic; anti-HIV;
KW Env; Gag; Pol; Nef; genetic immunisation; protein immunisation; AIDS;
KW acquired immunodeficiency syndrome; cytotoxic T lymphocyte; CTL.

XX
OS Synthetic.

XX
PN WO200232943-A2.

XX
PD 25-APR-2002.

XX
PF 14-AUG-2001; 2001WO-US25721.

XX
PR 14-AUG-2000; 2000US-225097P.

XX
PR 14-NOV-2000; 2000US-252115P.

XX
PR 28-MAR-2001; 2001US-279257P.

XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
(CHAD/) CHADRABARTI B K.

XX
PI Nabel GJ, Huang Y;
XX
DR WPI: 2002-452382/48.

XX
PT New modified human immunodeficiency virus Env, Gag, Pol of Nef DNA and
PT its encoded protein, useful as vaccines for genetic or protein
PT immunisation for acquired immunodeficiency syndrome or human
PT immunodeficiency virus infection -

XX
PS Disclosure; page 577-579; 794pp; English.

XX
XX The invention relates to a nucleic acid molecule encoding a modified HIV
CC (human immunodeficiency virus) protein. The nucleic acid molecule and its
CC encoded protein (modified HIV Env, Gag, Pol or Nef) are useful for
CC genetic or protein immunisation to a host, respectively. In particular
CC these are useful for ameliorating the symptoms of acquired
CC immunodeficiency syndrome (AIDS) or HIV infection and generating an
CC antibody or cytotoxic T lymphocyte (CTL) response against native HIV Env,
CC Gag, Pol or Nef. Furthermore, the nucleic acids, proteins and subsequent
CC compositions are useful for treating or preventing HIV infections or
CC AIDS. Sequences ABR91449-ABK91624 represent modified HIV protein-encoding
CC plasmid DNA sequences of the invention.

XX
SQ Sequence 7552 BP; 1797 A; 2149 C; 1995 G; 1611 T; 0 other:

Query Match 10.1%; Score 38; DB 24; Length 7552;
Best Local Similarity 62.8%; Pred. No. 0.78;
Matches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 26 TCCGCGCCCTTCTCTCCACCTTCCCGAGCCCTTAAGGGGCGGGGCCCAAGCCGAGG 85
DB 4066 TCCAGACCTCTGCTGCGCCGCGCAGCGGCGGCCGCCGCCCGAGGCGACGAGGAGG 4125
OY 86 GCGCTGCGCTTACCCGAGCGGAGGAGGCCCGCAG 119
DB 4126 GCGGCGAGCGGAGCGGAGCGGAGCGGAGCGGCGGCCAG 4159

RESULT 6

ABK70011/c
ID ABR70011 standard; DNA; 1111 BP.

XX
AC ABR70011;

XX
DT 15-JUL-2002 (first entry)

DE cDNA encoding human Pro peptide #51.

XX
KW Human; ss; gene; PRO; secreted protein; transmembrane protein;
KW genetic disorder; tumour; cancer.

XX
OS Homo sapiens.

XX
PN WO200224888-A2.

XX
PD 28-MAR-2002.

XX
PF 29-AUG-2001; 2001WO-US27099.

XX
PR 01-SEP-2000; 2000US-229896P.

XX
PR 05-SEP-2000; 2000US-230621P.

XX
PR 22-SEP-2000; 2000US-235147P.

XX
PR 10-NOV-2000; 2000WO-US30873.

XX
PR 12-JAN-2001; 2001US-261878P.

XX
PR 16-JAN-2001; 2001US-261910P.

XX
PR 16-JAN-2001; 2001US-262150P.

XX
PR 25-JAN-2001; 2001US-264395P.

XX
PR 02-FEB-2001; 2001US-266421P.

XX
PR 09-FEB-2001; 2001US-267623P.

XX
PR 28-FEB-2001; 2001WO-US06520.

XX
PR 09-MAR-2001; 2001US-274399P.

XX 23-JAN-2002 (first entry)
DT Human nervous system related polynucleotide SEQ ID NO 12346.
XX
DE
XX
XX Human; nocitropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX Homo sapiens.
OS
PN WO200159063-A2.
XX
XX
PD 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
PF
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-02094515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234424.
PR 21-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249269.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.

[illegible]

REFERENCE 2 (bases 1 to 81512)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 81512)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 81512)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 5 (bases 1 to 81512)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (28-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jan 28, 2003 this sequence version replaced gi:20376952.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.3% of Sequence;
 Estimated Total Number of Errors is 0.4.
 NOTE: This insert is not the entire sequence of the clone (entire sequence is 139.5b). It is clipped at the overlaps with AC008378 and AC106731. The number of bases overlapped with AC008378 is 4442 bps and with AC106731 is 15413 bps.

FEATURES
 source
 1. 81512
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CNC-281H14"
 /clone_11b="CHOR1-501"

BASE COUNT 20011 a 19534 c 19980 g 21987 t

ORIGIN

Query Match 96.5%; Score 362; DB 9; Length 81512;
 Best Local Similarity 99.2%; Pred. No. 1.3e-85;
 Matches 375; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 1 AGGCCCCCGGACCTATCTGCTCCGCGCCCTTCTCTCCACCTCCGAGACCCCTAA 60
 |||||
 DB 54172 AGGCCCCCGGACCTATCTGCTCCGCGCCCTTCTCTCCACCTCCGAGACCCCTAA 54113
 |||||

OY 61 AGGGGGGGGGGGCCCAAGCCGAGGGGCGCTGCGCTGACCCCGAGCGAAGGGCCCACT 120
 |||||
 DB 54112 AGGGGGGGGGGGCCCAAGCCGAGGGGCGCTGCGCTGACCCCGAGCGAAGGGCCCACT 54053
 |||||

OY 121 CTAGTCTTAATGCGGGTGCCTCTTTGACAGCGGGCGTTGGGACAAACAGCGGG 180
 |||||
 DB 54052 CTAGTCTTAATGCGGGTGCCTCTTTGACAGCGGGCGTTGGGACAAACAGCGGG 53993
 |||||

OY 181 ACGAGAGTAGTGACATACACAGAGATTTGGTGGCGGCGTGATCTCTCTCCG 240
 |||||
 DB 53992 ACGAGAGTAGTGACATACACAGAGATTTGGTGGCGGCGTGATCTCTCTCCG 53933
 |||||

OY 241 ACAGGAAACGGGAGCTATTTAAAGACCCATGATTAATCTTTCT--GAAA 297
 |||||
 DB 53932 ACAGGAAACGGGAGCTATTTAAAGACCCATGATTAATCTTTCTCGAGAGAA 53873
 |||||

OY 298 GCTTCTTGGGAGAGACAAAGATTTCCCTGCTTAAAGACCAAGGCGACACAGCGAG 357
 |||||
 DB 53872 GCTTCTTGGGAGAGACAAAGATTTCCCTGCTTAAAGACCAAGGCGACACAGCGAG 53813
 |||||

OY 358 GGTCGACAGGCGACGC 375
 |||||
 DB 53812 GGTCGACAGGCGACGC 53795
 |||||

RESULT 2
 AL669924/c 137345 bp DNA linear PRI 24-APR-2002
 LOCUS Human DNA sequence from clone XHbac-2L19 on chromosome 6, complete
 DEFINITION sequence.
 ACCESSION AL669924
 VERSION AL669924.9 GI:20068712
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
 Peck, A.
 Direct Submission
 JOURNAL Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Apr 7, 2002 this sequence version replaced gi:19572797.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPep; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep_xhbac-2L19_1s
 from a CHOR1-501 human bac - PGF cell line library VECTOR: PTARBAC2.1
 This sequence was generated from part of bacterial clone contigs constructed by the MHC Haplotype Consortium and collaborators. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr6/MHC.

FEATURES
 source
 1. 137345
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="XHbac-2L19"
 /clone_11b="CHOR1-501"
 /clone_11d="CHOR1-501"

BASE COUNT 33377 a 33207 c 33325 g 37436 t

ORIGIN

Query Match 96.5%; Score 362; DB 9; Length 137345;
 Best Local Similarity 99.2%; Pred. No. 1.4e-85;
 Matches 375; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 1 AGGCCCCCGGACCTATCTGCTCCGCGCCCTTCTCTCCACCTCCGAGACCCCTAA 60
 |||||
 DB 54398 AGGCCCCCGGACCTATCTGCTCCGCGCCCTTCTCTCCACCTCCGAGACCCCTAA 54339
 |||||

OY 61 AGGGGGGGGGGGCCCAAGCCGAGGGGCGCTGCGCTGACCCCGAGCGAAGGGCCCACT 120
 |||||
 DB 54338 AGGGGGGGGGGGCCCAAGCCGAGGGGCGCTGCGCTGACCCCGAGCGAAGGGCCCACT 54279
 |||||

OY 121 CTAGTCTTAATGCGGGTGCCTCTTTGACAGCGGGCGTTGGGACAAACAGCGGG 180
 |||||
 DB 54278 CTAGTCTTAATGCGGGTGCCTCTTTGACAGCGGGCGTTGGGACAAACAGCGGG 54219
 |||||

OY 181 ACGAGATTAAGTGACATACACAGAGATTTGGTGGCGGCGTGATCTCTCTCCG 240
 |||||

[illegible][illegible][illegible][illegible][illegible]


```
* 30902 31001: gap of unknown length
* 31002 32159: contig of 1158 bp in length
* 32160 32259: gap of unknown length
* 32260 33788: contig of 1529 bp in length
* 33789 33988: gap of unknown length
* 33889 35134: contig of 1246 bp in length
* 35135 35235: gap of unknown length
* 35235 36486: contig of 1252 bp in length
* 36487 36586: gap of unknown length
* 36587 38385: contig of 1799 bp in length
* 38386 38485: gap of unknown length
* 38486 39714: contig of 1229 bp in length
* 39715 39814: gap of unknown length
* 39815 41078: contig of 1264 bp in length
* 41079 41178: gap of unknown length
* 41179 42273: contig of 1095 bp in length
* 42274 42373: gap of unknown length
* 42374 44123: contig of 1750 bp in length
* 44124 44223: gap of unknown length
* 44224 45308: contig of 1685 bp in length
* 45309 46008: gap of unknown length
* 46009 47545: contig of 1537 bp in length
* 47546 50012: gap of unknown length
* 50013 50112: contig of 2367 bp in length
* 50113 51978: gap of unknown length
* 51979 52078: contig of 1866 bp in length
* 52079 54113: contig of 2035 bp in length
* 54114 54213: gap of unknown length
* 54214 55586: contig of 1373 bp in length
* 55587 55686: gap of unknown length
* 55687 57405: contig of 1719 bp in length
* 57406 57505: gap of unknown length
* 57506 58790: contig of 1285 bp in length
* 58791 58890: gap of unknown length
* 58891 60453: contig of 1563 bp in length
* 60454 60553: gap of unknown length
* 60554 62149: contig of 1596 bp in length
* 62150 62249: gap of unknown length
* 62250 63910: contig of 1661 bp in length
* 63911 64010: gap of unknown length
* 64011 65910: contig of 1900 bp in length
* 65911 66010: gap of unknown length
* 66011 67754: contig of 1744 bp in length
* 67755 67854: gap of unknown length
* 67855 69858: contig of 2005 bp in length
* 69859 69959: gap of unknown length
* 69960 71962: contig of 2003 bp in length
* 71963 72062: gap of unknown length
* 72063 74371: contig of 2309 bp in length
* 74372 74471: gap of unknown length
* 74472 76499: contig of 2028 bp in length
* 76500 76599: gap of unknown length
* 76600 78329: contig of 1730 bp in length
* 78330 78429: gap of unknown length
* 78430 81033: contig of 2604 bp in length
* 81034 81133: gap of unknown length
* 81134 82450: contig of 1317 bp in length
* 82451 82550: gap of unknown length
* 82551 85880: contig of 3330 bp in length
* 85881 85980: gap of unknown length
* 85981 89959: contig of 3979 bp in length
* 89960 92499: gap of unknown length
* 92500 92599: contig of 2440 bp in length
* 92600 95372: gap of unknown length
* 95373 95472: gap of unknown length
* 95473 98796: contig of 3324 bp in length
* 98797 98896: gap of unknown length
* 98897 101619: contig of 2723 bp in length
* 101620 101719: gap of unknown length
* 101720 105463: contig of 3744 bp in length
* 105464 105563: gap of unknown length
```

```
* 105564 108520: contig of 2957 bp in length
* 108521 108620: gap of unknown length
* 108621 111714: contig of 3094 bp in length
* 111715 111814: gap of unknown length
* 111815 115701: contig of 3887 bp in length
* 115702 115801: gap of unknown length
* 115802 117720: contig of 1919 bp in length
* 117721 117820: gap of unknown length
* 117821 121379: contig of 3459 bp in length
* 121380 121480: gap of unknown length
* 121481 125489: contig of 4110 bp in length
* 125490 125897: gap of unknown length
* 125898 128997: contig of 3308 bp in length
* 128999 133134: gap of unknown length
* 133135 133234: contig of 4137 bp in length
* 133235 137402: gap of unknown length
* 137403 137503: gap of unknown length
* 137504 142642: contig of 5140 bp in length
* 142643 142742: gap of unknown length
* 142743 148552: contig of 5810 bp in length
* 148553 148653: gap of unknown length
* 148654 158806: contig of 10154 bp in length.
```

FEATURES

source

1. 158806

Location/Qualifiers

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="Y/NK"

/clone="RP24-335114"

misc-feature

/note="assembly_name:Contig10"

misc-feature

/note="assembly_name:Contig22"

misc-feature

/note="assembly_name:Contig34"

Query Match 61.1%; Score 229; DB 2; Length 158806;
Best Local Similarity 83.2%; Pred. No. 4.6e-50; Indels 13; Gaps 4;
Matches 312; Conservative 0; Mismatches 50;

```
QY 6 CCGCGGACCCATCTTCCTGCGCGCCCTTCCTCCAGCCCTCCGAGCCCTAAAGGG 65
DB 138522 CCGGAACTTTCATGCGGCGGCTGCGATCTTCCTCCAGCCCTCCGAGAGCGAA----- 138576
QY 66 CCGCGGCGCCCAAGCCGAGGCGCTGCGCTGACCCCGAGCGGAGGCCCACTAGG 125
DB 138577 ---GGACAGCGGAGCCGAGGCGCGCGGTGACCCGAGGAGGAGGCCCACTAGG 138633
QY 126 TCTTAATGCGGCTGCTCTTTCGACAGGCGGCTTGGGAGCAACAGCGGAGGAG 185
DB 138634 TCTTAATGCGGCTGCTCTTTCGACAGGCGGCTTGGGAGCAACAGCGGAGGAG 138693
QY 186 AGATAGGTGATATACAGAGCAGATTTGGCGCGCTGATACCTCTCCGACAGG 245
DB 138694 AGATAGGTGATATACAGAGCAGATTTGGCGCGCTGATACCTCTCCGACAGG 138753
QY 246 AAACGCGAGCTATT---AAAGACCTATGCACTTATCTTCC---TGAAAGCT 301
DB 138754 AAACGCGAGCTATTAAAGGCGCTATGCACTTATCTTCCCGAGGAAACTT 138813
QY 302 CTTCGCGAGGAGCAAAAGATGT-TCCCGGCTTAAGACCAAGGCGACACAGAGGAGT 360
DB 138814 CTTCGCGAGGAGCAAAAGATGTCTCCCTTAAAGATACCAAGGCGACAGAGGAGT 138873
QY 361 CTGCACAGCGAGCG 375
DB 138874 TTGTACAGCGAGCG 138888
```

RESULT 6
CNS08CBD/c
LOCUS CNS08CBD 211509 bp DNA linear PRI 05-SEP-2002

[illegible]

Query Match	Similarity	Score	DB	Length	21/09/21
Matches	94: Conservation	or Mismatches	wt	Index	92
05	7	1000	1000	1000	1000
06	98775	1000	1000	1000	1000
07	57	1000	1000	1000	1000
08	98715	1000	1000	1000	1000
09	127	1000	1000	1000	1000
10	98655	1000	1000	1000	1000

```

DEFINITION Homo sapiens zinc finger transcriptional regulator (GOS24) gene,
complete cds.
ACCESSION M92844
VERSION M92844.1 GI:2072389
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 3889)
AUTHORS Blum, S., Forsdyke, R.E. and Forsdyke, D.R.
TITLE Three human homologs of a murine gene encoding an inhibitor of stem
cell proliferation
JOURNAL DNA Cell Biol. 9 (8), 589-602 (1990)
MEDLINE 91103879
PUBMED 2271120
REFERENCE 2 (sites)
AUTHORS Taylor, G.A., Lai, W.S., Oakley, R.J., Seidlin, M.F., Shows, T.B.,
Eddy, R.L. Jr. and Blackshear, P.J.
TITLE The human TTP protein: sequence, alignment with related proteins,
and chromosomal localization of the mouse and human genes
JOURNAL Nucleic Acids Res. 19 (12), 3454 (1991)
MEDLINE 91288233
PUBMED 2062660
REFERENCE 3 (bases 1 to 3889)
AUTHORS Heximer, S.P. and Forsdyke, D.R.
TITLE A human putative lymphocyte G0/G1 switch gene homologous to a
rodent gene encoding a zinc-binding potential transcription factor
JOURNAL DNA Cell Biol. 12 (1), 73-88 (1993)
MEDLINE 93135830
PUBMED 8422274
REFERENCE 4 (bases 1 to 3889)
AUTHORS Heximer, S.P., Cristillo, A.D., Russell, L. and Forsdyke, D.R.
TITLE RT-PCR analysis of RNA of the CCH zinc finger protein encoding
gene GOS24 (TIS11/TTP/NUP475) in cultured human blood mononuclear
cells
JOURNAL Unpublished
COMMENT On May 6, 1997 this sequence version replaced gi:183444.
FEATURES
Source
location/Qualifiers
1..3889
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="19q13.1"
/tissue_type="placenta"
549..630
/gene="GOS24"
/exon
/number=1
/evidence=experimental
join(607..630,1441..2397)
/gene="GOS24"
CDS
join(607..630,1441..2397)
/gene="GOS24"
/exon
/codon_start=1
/product="zinc finger transcriptional regulator"
/protein_id="FAC37600.1"
/db_xref="GI:183445"
/translation="MDLTAIESLISLSPVVPVPSDHGTESSFGWSSGFGWISPSD
SSPGVTSRLPGSTSLVEGSCGWPVPPGFAPLAPRLGTEISPSPTATSTPS
RYKELCRTSESRCRYGAKCFAHGELRQANRRHPKYTELCHRFYLOGRCPTGS
RCHFTIHPMEDILAPGHPVPIROSISFSGLPSPGGLAGPLSSSSFSFSS
PPPPGDLPLSFAFSAAGTFLARDPPTVCCPCSCRRATPTISWAGPLGGIARTSYVS
LGSDDPEYASSGSSLSGSDSPFEAGVFAPOVPAAPARLPIFNRTISVE"
1441..3103
/gene="GOS24"
/exon
/number=2
/evidence=experimental
2978..2986
/feature="TA-rich conserved element (TARCE)"
3163..3168
/misc_feature
/note="U-rich RNA polymerase II termination element"
3198..3889
/misc_feature
/note="Cpg island"

```

```

BASE COUNT 648 a 1304 c 1078 g 850 t 9 others
ORIGIN chromosome 19q13.1.
Query Match 11.7%; Score 43.8; DB 9; Length 3889;
Best Local Similarity 59.1%; Pred. No. 0.73;
Matches 75; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 2 GGGCCCCGACCCATCCGCTCCGCGCTTCCTTCACACCCCTCCGACCCCTTAA 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 GGGCAGGCGTCCCGCATCCGACCCGACCCCTTCGCCGACATTCGCCGCTCAG 360
62 GGGCGCGCGGCGCCAGCGCGCTCCGCTGACCCCGAGGGAAGGCCCTCAGTC 121
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GCGTGTCCACCGCAGGCTCAGGCGCTCCGCGCGCGGCGGAGGGAACCACTG 420
QY 122 TAGGTCC 128
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 CAGGGCC 427

```

```

RESULT 8
AC058782 171852 bp DNA linear HTG 22-FEB-2001
LOCUS Homo sapiens chromosome 1 clone RP11-285F7, WORKING DRAFT SEQUENCE,
8 unordered pieces.
DEFINITION
ACCESSION AC058782
VERSION AC058782.11 GI:12331477
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 171852)
AUTHORS Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,
Fedorpiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
Mao, J., Marathe, R., Morehouse, A.J., Oefner, P., Palm, C.J.,
Ramirez, D., Wilhelm, J., Yu, S. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT On Jan 22, 2001 this sequence version replaced gi:12203632.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDSTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: RP11-285F7
Center clone name: RP11-285F7
----- Summary Statistics
Sequencing Vector: M13mp18, X02513
Chemistry: Dye-terminator 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165994 bases at least Q40
Consensus quality: 167611 bases at least Q30
Insert size: 165696; agarose-fp
Insert size: 171152; sum-of-contigs
Quality coverage: 9.4x in Q20 bases; agarose-fp
Quality coverage: 9.1x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```



```
* 88572 88671: gap of 100 bp
* 88672 102917: contig of 14246 bp in length
* 102918 103017: gap of 100 bp
* 103018 120081: contig of 17064 bp in length
* 120082 120181: gap of 100 bp
* 120182 136057: contig of 15876 bp in length
* 136058 136157: gap of 100 bp
* 136158 156302: contig of 20145 bp in length
* 156303 156402: gap of 100 bp
* 156403 167787: contig of 11385 bp in length
* 167788 167888: gap of 100 bp
* 167889 183109: contig of 15222 bp in length
* 183110 196606: contig of 13397 bp in length
* 196607 196706: gap of 100 bp
* 196707 218924: contig of 22218 bp in length
* 218925 219024: gap of 100 bp
* 219025 249572: contig of 30548 bp in length
* 249573 249672: gap of 100 bp
* 249673 287188: contig of 37516 bp in length
* 287189 287288: gap of 100 bp
* 287289 289388: contig of 2100 bp in length
* 289389 289488: gap of 100 bp
* 289489 291859: contig of 2371 bp in length
* 291860 291959: gap of 100 bp
* 291960 294128: contig of 2169 bp in length
* 294129 296882: gap of 100 bp
* 296883 296982: gap of 100 bp
* 296983 299603: contig of 2621 bp in length
* 299604 301793: gap of 100 bp
* 301794 301893: gap of 100 bp
* 301894 305076: contig of 3183 bp in length
* 305077 305176: gap of 100 bp
* 305177 308174: contig of 2998 bp in length
* 308175 308274: gap of 100 bp
* 308275 310542: contig of 2268 bp in length
* 310543 310642: gap of 100 bp
* 310643 313870: contig of 3228 bp in length
* 313871 313970: gap of 100 bp
* 313971 317124: contig of 3154 bp in length
* 317125 317224: gap of 100 bp
* 317225 320595: contig of 3371 bp in length
* 320596 320695: gap of 100 bp
* 320696 323155: contig of 2460 bp in length
* 323156 323255: gap of 100 bp
* 323256 326284: contig of 3029 bp in length
* 326285 326384: gap of 100 bp
* 326385 329626: contig of 3242 bp in length
* 329627 329726: gap of 100 bp
* 329727 333482: contig of 3756 bp in length
* 333483 333582: gap of 100 bp
* 333583 337189: contig of 3607 bp in length
* 337190 337289: gap of 100 bp
* 337290 341875: contig of 4586 bp in length
* 341876 341975: gap of 100 bp
* 341976 344388: contig of 2413 bp in length
* 344389 344488: gap of 100 bp
* 344489 349154: contig of 4666 bp in length
* 349155 349254: gap of 100 bp
* 349255 353824: contig of 4570 bp in length
* 353825 353924: gap of 100 bp
* 353925 357953: contig of 4029 bp in length
* 357954 358053: gap of 100 bp
* 358054 363695: contig of 5642 bp in length
* 363696 363795: gap of 100 bp
* 363796 367278: contig of 3483 bp in length
* 367279 367378: gap of 100 bp
* 367379 371902: contig of 4524 bp in length
* 371903 372002: gap of 100 bp
* 372003 377117: contig of 5115 bp in length
* 377118 377217: gap of 100 bp
```

```
* 377218 382171: contig of 4954 bp in length
* 382172 382271: gap of 100 bp
* 382272 388633: contig of 6562 bp in length
* 388634 388933: gap of 100 bp
* 388934 393530: contig of 4597 bp in length
* 393531 393630: gap of 100 bp
* 393631 398934: contig of 5304 bp in length
* 398935 399035: gap of 100 bp
* 399036 404551: contig of 5517 bp in length
* 404552 404651: gap of 100 bp
* 404652 410493: contig of 5842 bp in length
* 410494 410593: gap of 100 bp
* 410594 415944: contig of 5351 bp in length
* 415945 416044: gap of 100 bp
* 416045 424416: contig of 8372 bp in length
* 424417 424516: gap of 100 bp
* 424517 430287: contig of 5771 bp in length
* 430288 430387: gap of 100 bp
* 430388 437707: contig of 7320 bp in length
* 437708 437807: gap of 100 bp
* 437808 445871: contig of 8064 bp in length
* 445872 445971: gap of 100 bp
* 445972 455450: contig of 9479 bp in length.
```

FEATURES

```
source
1..455450
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="11"
  /map="11"
  /clone="MGS1-213E3"
  /clone_1lb="genome Systems Mouse BAC 1"
```

BASE COUNT 119379 a 104363 c 107639 g 117130 t 6939 others

ORIGIN

Query Match 11.2%; Score 42; DB 2; Length 110000;
Best Local Similarity 56.5%; Pred. No. 3.5;
Matches 78; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
QY 5 CCCCCGACCCCTATCTGCTCCCGCCCTTCTCTGACACCTCCCGGACCCCTAAAGG 64
   || || || || || || || || || || || || || || || || || || || ||
Db 72293 CTTCCCTCCCTCTCTCTCCGCGCGCGCCCTCCCTCCCGGGGGCCCGGAGCG 72352

QY 65 GCGCGGGGGCCCAAGCGAGGCGCGCTGACCCCGAGCGAGGAGCCCGAGCTAG 124
   || || || || || || || || || || || || || || || || || || || ||
Db 72353 GGGGGAGGAGCGCGCGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 72412

QY 125 GTCTATGCGGCGTGGCG 142
   || || || || || || || || || || || || || || || || || || || ||
Db 72413 GCGCGGGGGGGGGGGCG 72430
```

RESULT 10

AC026512/c

LOCUS

Homo sapiens chromosome 15, clone Rpl1-37509, complete sequence.

AC026512

AC026512.6

GI:15983569

HTG.

Homo sapiens (human)

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 90707) Nusbaum, C., Lander, E., Abraham, H., Allen, N., Birren, B., Linton, L., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,


```

repeat_region /rpt_family="L2"
complement(4926..5672)
/rpt_family="L2"
repeat_region complement(6570..6829)
/rpt_family="MIR"
repeat_region 7457..7493
/rpt_family="(TTTAA)n"
complement(7884..8219)
/rpt_family="L2"
repeat_region 8517..8554
/rpt_family="(CA)n"
repeat_region 8996..9563
/rpt_family="MER90"
repeat_region 9649..9718
/rpt_family="MER5B"
repeat_region 10054..10253
/rpt_family="L2"
repeat_region 10325..10440
/rpt_family="L2"
repeat_region 10610..10822
/rpt_family="MIR"
complement(10881..11030)
/rpt_family="MER5A"
repeat_region 11046..11146
/rpt_family="L2"
repeat_region 11983..12096
/rpt_family="MER5A"
complement(12275..12452)
/rpt_family="MER5A"
complement(13065..13238)
/rpt_family="Charlie8"
repeat_region 13749..13788
/rpt_family="MLT1H1"
repeat_region 13789..14092
/rpt_family="Alusp"

```

Query Match 11.1%; Score 41.8; DB 9; Length 90707;

Best Local Similarity 49.3%; Pred. No. 3.9; Matches 109; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```

OY 10 GCACCCCTCATCTGGCTCCGCCCTTCTCTCCACCTCCCGGACCCCTAAAGGGGGC 69
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 68040 GCGCGCCGCGCGGATCCACCTGACCGCCTCTCGCCCGGACCTCCAAACCCGCC 67981
OY 70 GGGGCCCAAGCGGAGGGGCTGCGGCGGACCCCGAGGGAAGGGCCAGTCTAGTCT 129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67980 AAGCTTGGCGCCACAGGAGCGCGCGGACGCCCCAGAGCCGCCCGTCAAGAGGCC 67921
OY 130 AATCGGGGTGCGTCTCTTGGACAGCGCGGCTTGGGGACACAGCGGGGAGAGAT 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67920 GCGTCTCGGCTCTCGCTCTCCCGTCCCGACCGCTCTCAAGGAGCGCGCTCC 67861
OY 190 AAGGTGACATACGACAGCAGATTTGGTGGCGGCGCTGATAC 230
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67860 GGGCTTCTTAACCTGACGAGCGCTGGAGGGGAGCAGACAC 67820

```

RESULT 11

AL773527 137286 bp DNA linear MAM 07-SEP-2002
 LOCUS AL773527 Pig DNA sequence from clone XX-707FL, complete sequence.
 DEFINITION AL773527
 ACCESSION AL773527 GI:21998261
 VERSION
 KEYWORDS HTG.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 137286)
 AUTHORS Sehra,H.
 TITLE Direct Submission
 JOURNAL Submitted (06-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 26, 2002 this sequence version replaced gi:21912758.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information on the WormPeP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-707FL is from a pig library
 VECTOR: pBeloBAC11.
 Location/Qualifiers
 source 1..137286
 /organism="Sus scrofa"
 /mol_type="genomic DNA"
 /db_xref="taxon:9823"
 /clone="XX-707FL"
 /clone_lib="SBAB"

BASE COUNT 32551 a 35854 c 35765 g 33116 t

ORIGIN

Query Match 11.1%; Score 41.6; DB 4; Length 137286;

Best Local Similarity 50.5%; Pred. No. 4.7; Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```

OY 25 CTCGGCGCCCTTCTCTCACCTTCCGACCCCTTAAGGGGCGGGGCCCAAGCCGAG 84
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 63759 CACACGGCAGCAGCGCGCCGCCCTTCTTACGCCACGACGCGGCGCGCG 63818
OY 85 GCGCTGCGCTTACCCCGAGGAGGCGCCCAAGTCTAGTCTCTAATGGCGGTGGCGTC 144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 63819 GCGCGCCCGCTGACACCCCGCGCTGCGCTCCCAAGGCTTGGCCGTGTGTGGGCG 63878
OY 145 TCTTTGACAGCGGCGGTTTGGGACACAGCGGAGAGAGATTAAGTACATACAG 204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 63879 GCAGCGGCGCCCGGGGGGTGGGGAATAGGGGGTTAAAAAAGCGCGGAAACCGG 63938
OY 205 AGCAGATTGGTGGCGCGCGC 224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 63939 GCGAGGTGAGAGCGCGCTGC 63958

```

RESULT 12

OSJN00209/c 164907 bp DNA linear PLN 17-JUL-2002
 LOCUS OSJN00209/c Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNB0058J09,
 DEFINITION complete sequence.
 ACCESSION AL663008
 VERSION AL663008.2 GI:21742449
 KEYWORDS HTG.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Liliopsida; Poales; Poaceae;
 Euphorbiales; Oryzaceae; Oryza.
 REFERENCE 1

